Dynamic Bayesian Networks for the Classification of Spinning Discs

by

Aurora Clare Schmidt

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2004

Author .....
Department of Electrical Engineering and Computer Science
May 20, 2004

Certified by .....
John W. Fisher III
Principal Research Scientist
MIT Thesis Supervisor

Certified by .....
David P. Cebula
MIT Lincoln Laboratory
VI-A Company Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Dynamic Bayesian Networks for the Classification of Spinning Discs

by

Aurora Clare Schmidt

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2004, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Computer Science and Engineering

Abstract

This thesis considers issues for the application of particle filters to a class of nonlinear filtering and classification problems. Specifically, we study a prototype system of spinning discs. The system combines linear dynamics describing rotation with a nonlinear observation model determined by the disc pattern, which is parameterized by angle. A consequence of the nonlinear observation model is that the posterior state distribution of angle and spin-rate is multi-modal. This detail motivates the use of particle filtering. Practical issues that we consider when using particle filters are sample depletion and sample degeneracy, both of which lead to poor representations of the state distributions.

Variance based resampling and regularization are common methods to mitigate sampling issues in particle filtering. We investigate these methods empirically for our prototype problem. Specific parameters of interest relating to these methods are the number of particles used to approximate the posterior distribution, quantitative methods for deciding when to resample, choice of regularization variance, the impact of measurement noise on all of these, and performance over time.

A common issue, leading to inaccurate sample-based representations, is the case of relatively low measurement noise combined with an insufficient number of particles. Our empirical results show that for relatively smooth patterns (e.g. linear, cosine) particle filters were less susceptible to sampling issues than for patterns with higher frequency content. The goal of our experiments is to quantify the nature of these differences.

MIT Thesis Supervisor: John W. Fisher III
Title: Principal Research Scientist

VI-A Company Supervisor: David P. Cebula
Title: MIT Lincoln Laboratory

Acknowledgments

I would like to thank my thesis advisors David Cebula, John Fisher, and Keh-Ping Dunn. Your exceptional guidance and patience was greatly appreciated. I would also like to thank and credit James R. Williamson for his great ideas and excellent discussions. The advice of Jason L. Williams was also key in understanding the limitations of the particle filter and developing methods to test and compare the performance. Thanks to Dan Rudoy, Christi Electris, David Choi and all the members of Lincoln Laboratory's Group 32 for your support and advice. Finally, thanks to John Monaco, who provided the advanced computational technology necessary for this research.

Contents

Chapter 1

Introduction

This thesis studies the use of Particle Filtering as a tool for approximate inference in Dynamic Bayesian Networks. The objective is to estimate the state of a spinning disc system from sequential observations, where the mappings between states and observations are not linear. While the conditions are similar to traditional Kalman Filtering problems, the fundamental difference is that the nonlinear mapping of states to observations can result in multi-modal posterior state distributions. Multi-modal distributions are not handled by traditional Kalman Filtering Techniques nor extensions of these techniques.

The objective is to evaluate the the performance of Particle Filtering for the estimation and classification of unknown objects. The main advantage expected from this stochastic sampling technique, is that multi-modal probability spaces can be modeled. However, disadvantages must arise from the inexact nature of the algorithm as well as the computational costs of sampling. As a result, we investigate several techniques to increase the efficiency of the algorithm, including regularization and controlled resampling.

Chapter 2 will cover background information on Dynamic Bayesian Networks, Hidden Markov Models, and inference using a sequence of observations.

Chapter 3 discusses the motivations behind the Particle Filtering technique as well as the basic algorithm.

Chapter 4 addresses the main issues of Particle Filtering, including sample de-

generacy and sampling depletion. The section goes on to describe commonly used techniques to mitigate these problems.

Chapter 5 introduces the Spinning Discs scenario, which is the simple system for which Particle Filtering will be tested as an estimator and classifier.

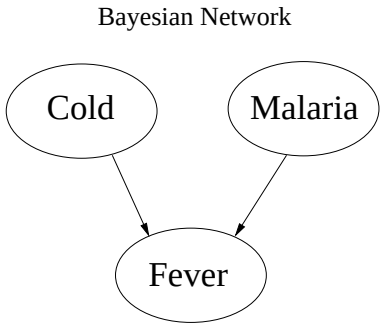
Chapter 6 presents the results of tests on various patterned discs, including the linear and cosine patterns. In addition, this section demonstrates discrimination between more challenging nonlinear patterns and evaluates the particle filter's performance as a classifier.

Finally, Chapter 7 will discuss lessons learned and ideas for future work.

Chapter 2

Background

Learning with bayesian networks is a growing field because of the meaningful structure of conditional dependency networks and the grounding of these models in probability theory. A bayesian network is a directed acyclic graph modeling the conditional dependence between variables. Each variable contains a conditional probability table for all values of its parents. In that way, any conditional distribution is described by the bayesian network model. Please see figure 2-1 for an example bayesian network.



Conditional Probability Table

C	M	P(F|C,M)
0	0	0.95
0	1	0.67
1	0	0.45
1	1	0.07

Figure 2-1: Example Bayesian Network and Conditional Probability Table

2.1 Bayesian Network Inference

Bayesian networks are excellent at capturing the causal relationships within a system. However, they are most useful when they can also infer probability distributions from evidence. The difficulty of inference is related to the complexity of the network.

Networks with many variables, states per variable, or dependencies between variables make probabilistic inference more challenging. The general rule for exact inference must involve the product of the probability of all variables given its parents, summed over all the states of every non-evidence variables. Let S be the set of all variables.

$$P(A|B,C) = \sum_{allstates} \left(\prod_{V \in S} P(V|Parents(V)) \right) \tag{2.1}$$

From the formula, it is clear that exact inference becomes increasingly difficult in large networks with many variables and dependencies. In addition, variables may have continuous values, which can further complicate inference. As a result, exact inference is intractable in many of the practical networks we may encounter.

2.2 Dynamic Bayesian Networks & Hidden Markov Models

Dynamic Bayesian Networks (DBN's) and Hidden Markov Models (HMM's) are common examples of large networks for which exact inference is difficult. A Dynamic Bayesian Network is a Bayesian Network with variables whose values change over time. In order to show this, the structure of the network is drawn at each time step as well as the dependencies of variables from one time-step to another. In practical systems variables often depend only on previous time-steps, but the Dynamic Bayesian Network places no restriction on how many other time-steps influence the current one, and whether anti-causal relationships exist.

One of the most common types of Dynamic Bayesian Network is the Hidden Markov Model. In a Hidden Markov Model we make the assumptions that the structure of each time-slice, comprised of the variables and links, does not change and that variables at each time depend only on variables in the current or previous time-slice. These two conditions describe a first-order Markov process. In addition, we also assume that observed variables must depend only on current state variables. For an example of a Hidden Markov Model, please see figure 2-2.

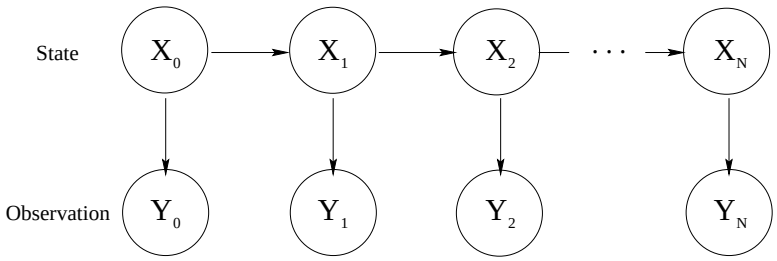


Figure 2-2: Example Hidden Markov Model. Each variable belongs to the time-slice denoted by its subscript. The X's are state variables, while Y's are observed variables.

According to Russell and Norvig's text, the major difference between Hidden Markov Models and Dynamic Bayesian Networks is that HMM's collect all the unobserved, or state, variables into one multi-dimensional variable. [15, p. 559] A key advantage of this simplification is that we can describe the system simply with a transition and measurement equation. The transition equation describes how the current state variables, X_t, depend on previous state variables, X_{t-1}. This equation is often made non-deterministic by the addition of extra terms representing system noise. The measurement equation relates the current observations, Y_t, to the current state, X_t, and can also include terms for measurement noise. The goal of a filter, or estimator, is to derive the probability distribution, $P(X_n|Y_1, Y_2, ..., Y_n)$, from the sequence of observations[1]. There is no limit to the length of this network. It grows with the length of the sequence of observations, making exact inference intractable using equation (2.1).

2.3 Kalman Filtering

However exact inference is still possible in cases where the transition and measurement equations are linear and the noise is additive and gaussian. The optimal [2] and efficient estimate is acheived by the well-known Kalman-Bucy Filter.[10, 11] With these simplifications the transition equation reduces to a linear combination, denoted by the matrix A, of the previous state plus zero-mean white gaussian noise, $\mathbf{p}$, with

[1] The sequence of observations is often denoted $Y_{1:n}$.

[2] *Optimal* refers to the filter that minimizes the sum of the squared errors.

covariance Q.

$$\mathbf{x_{t+1}} = A\mathbf{x_t} + \mathbf{p} \tag{2.2}$$

The measurement equation becomes a linear combination, denoted by the matrix H, of the current state plus noise, $\mathbf{v}$. We will consider v to be zero-mean, white, gaussian, and uncorrelated with $\mathbf{p}$ with the covariance matrix, R.

$$\mathbf{y_t} = H\mathbf{x_t} + \mathbf{v} \tag{2.3}$$

Under these circumstances, the Kalman Filter can re-estimate the least squares value of $\mathbf{x_t}$ and the covariance of this estimate at every time-step. This is possible since the transition and measurement models are linear and the noise is independent, additive, and gaussian. As a result, the posterior distribution will be gaussian and so can be described completely by its mean and covariance. In addition, there is little argument about what is the best estimate for the Kalman Filter, since the mean, mode, and median are all equal.[13]

A new estimate for $\mathbf{x_t}$, given the previous estimate, $\mathbf{x_{t-1}}$, and the covariance of the error for that estimate, P_t^-, are given by the Time Update Equations.

$$\mathbf{x_t^-} = A\mathbf{x_{t-1}} \tag{2.4}$$

$$P_t^- = AP_{t-1}A^T + Q \tag{2.5}$$

The new estimate for $\mathbf{x_t}$ using the observation, $\mathbf{y_t}$, and the covariance for the new estimate, P_t, are given by the Measurement Update Equations.

$$\mathbf{x_t} = \mathbf{x_t^-} + K(\mathbf{y_t} - H\mathbf{x_t^-}) \tag{2.6}$$

$$\text{where } K = P_t^- H^T (HP_t^- H^T + R)^{-1} \tag{2.7}$$

$$P_t = (I - KH)P_t^- \tag{2.8}$$

The Kalman Filter is often used for approximate inference in nonlinear filtering

problems. These are cases in which the transition or measurement model is no longer linear. An early nonlinear version of the Kalman Filter was called the Extended Kalman Filter (EKF). The EKF approximates the transition model as linear using the first order Taylor series expansion of the nonlinear equations.[7] This approach still assumes the posterior distribution will be similar to gaussian and therefore has only one mode. Although gaussianity is not preserved in nonlinear transformations, we assume that a single mode will still exist and that the estimate can again be modeled by a mean and covariance.

A more recent extension of the Kalman Filter is called the Unscented Kalman Filter (UKF) or the Sigma Point Kalman Filter (SPKF). Julier and Uhlmann pointed out that since we are assuming near gaussian posterior distributions, we can do better by modeling the nonlinear transformations of the sigma-points of the gaussian, rather than linearizing nonlinear equations.[9] The first and second order sigma points, the mean and covariance, are transformed using the nonlinear equations and have been shown to be a better model of the posterior distributions.[8, 16, 17] However, this approach again assumes posterior distributions with a single mode.

Unfortunately, in nonlinear problems, there is nothing to dictate that the posterior probability distribution must be unimodal. For example, a simple nonlinear measurement model can result in a multi-modal posterior density for the state given an observation. When there are multiple states that can explain a single observation, inference from that observation should result in a posterior density with more than one hypothesis. Figure 2-3 is an example measurement model that has multiple states, x, that account for a particular measurement, y.

For the class of problems where the measurements are not linear functions of the state, it is clear that an inference method that can model multiple hypotheses is needed. This is the motivation for this study of Particle Filtering. Particle Filtering is an approximate inference method that will attempt to sample the high probability regions of the posterior distribution. This technique will be applied to the example application of Spinning Discs in order to judge the considerations and efficiency of this inference algorithm. Chapter 3 will discuss the main ideas behind Particle Filtering

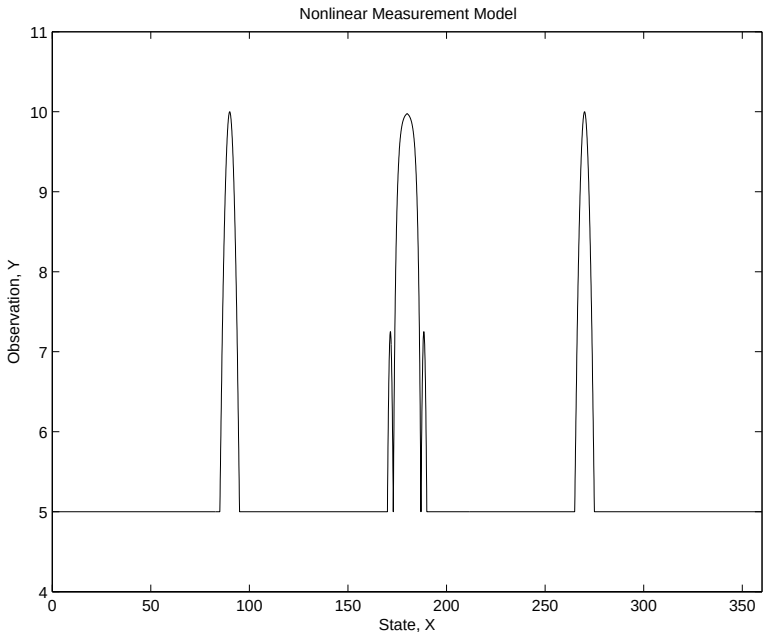


Figure 2-3: Example Nonlinear Measurement Model. Since there are multiple states that can explain a single observation, the inferred probability distribution for the state, X, given the observation, Y, may have many modes.

and outline the algorithm.

Chapter 3

Motivations for Particle Filtering

The traditional methods for nonlinear inference using sequential observations, known as the Extended and Unscented Kalman Filters, are not sufficient for problems with multi-modal posterior distributions.[1] Particle Filtering is an approximate inference technique which uses Monte-Carlo sampling to model the posterior distribution. The goal of this study is to use Particle Filtering for the estimation and classification of spinning discs and evaluate the advantages and considerations of this technique. In the following sections, the main ideas behind Particle Filtering will be outlined, as well as the basic algorithm.

The goal of Particle Filtering is to sample the posterior distribution, $P(X_n|Y_{1:n})$. Particle Filtering embodies two main sampling ideas, which are importance sampling and efficient sampling.

3.1 Importance Sampling

Importance sampling is one of the fundamental ideas of Particle Filtering. The idea is to model a desired distribution by re-weighting samples from another distribution, called the *proposal* distribution. The proposal distribution can be chosen to be anything, but should bear as much similarity to the desired distribution as possible. The prior distribution is often used as the proposal distribution in particle filtering.

[1] Please see section 2.3 for the discussion of Kalman Filtering Techniques.

After samples are generated according to the proposal distribution, a particle filtering algorithm re-weights those samples by the probability of the observations given each sample, $P(Y|x_s)$. After renormalization, the re-weighted samples will have relative weights according to the posterior distribution we wish to model, $P(X|Y)$. Let the proposal distribution be $P(X)$, the prior, for the purposes of illustration.

Importance Sampling

- Initial samples are sampled according to $P(X)$
- Weights are calculated by multiplying each original weight by $P(Y|x_s)$
- The new weights, $w = P(Y|x_s)P(x_s) = P(x_s|Y)P(Y)$, by Bayes Rule
- Renormalize the re-weighted samples since all contain the same $P(Y)$ term, and we have weighted samples of the posterior distribution, $P(X|Y)$.

$$P(X|Y_{1:n}) = P(X|Y_{1:n-1}) \left(\frac{P(Y_n|X)}{P(Y)} \right) \tag{3.1}$$

For a sequence of observations, the weights of each sample can be updated sequentially, as shown in equation 3.1. At each time-step the weight is multiplied by the probability of the new observation given the state. When renormalized, the weights represent the probability of each state given all the observations, $Y_{1:n}$. The reason why we can update the weights at each time-steps is that this is a Markov process, so the observations are conditionally independent given the state.

3.2 Efficient Sampling

The second main idea of Particle Filtering is that the whole distribution need not be sampled. Usually we are most interested in the regions of highest probability. Therefore it is more efficient to concentrate our samples in those regions.

Figure 3-1 demonstrates the idea of efficiency of sampling. The top graph shows relatively uniform sampling, while the bottom shows samples concentrated near areas of high probability. If we sampled the distribution in the figure more finely, and then resampled from those those samples, each of whose probability would be deter-

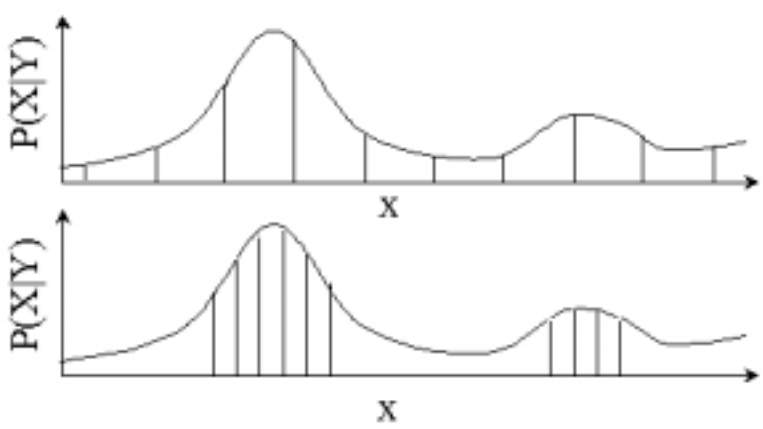


Figure 3-1: Example of More Efficient Sampling of a Distribution. The top graph shows relatively uniform sampling, while the bottom shows samples concentrated near areas of high probability.

mined by the distribution, we could create the lower graph with more samples in the high probability areas. The method of resampling from the distribution created by importance sampling concentrates samples near the peaks of the distribution.

Importance sampling when combined with resampling can be shown to converge on the posterior distribution.[4] Importance sampling allows the algorithm to model the posterior distribution with samples from another distribution, while resampling helps to model just the highest probability areas of the posterior distribution.

3.3 The Basic Algorithm

The basic particle filtering algorithm involves sampling with a proposal distribution, re-weighting those samples by the probability of the evidence given those states, and then resampling from the weighted samples.

Figure 3-2 is a graphical representation of the particle filtering algorithm. The first row of circles at the top of the figure are samples, also called *particles*, taken using a proposal distribution. In this case, the proposal distribution was a uniform prior. The initial samples start with equal weight. In order to weight the samples by their probability given the evidence, each sample weight is multiplied by the probability of the observation given the sample state, after which all the sample weights are

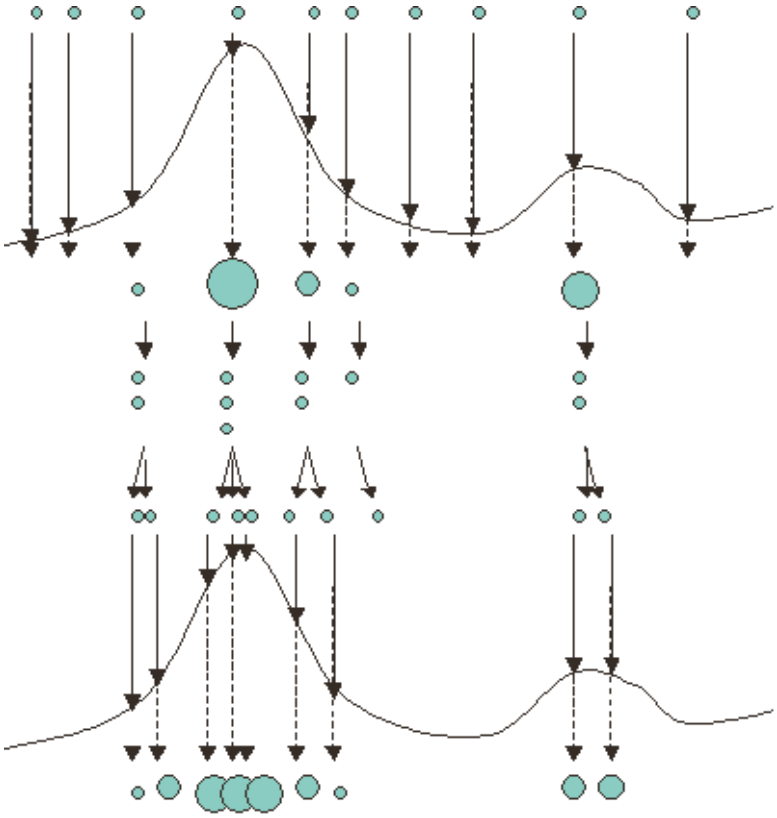


Figure 3-2: The Particle Filtering Algorithm. Each circle represents a sample particle with radius proportional to its weight. The algorithm proceeds from top to bottom. *similar figure in Sequential Monte Carlo Methods in Practice[5]*

renormalized. The second layer of circles, with differing weights, is the result of the Importance Sampling Step.

Next, for the resampling step, the algorithm stochastically samples the distribution created by the weighted samples we have generated. Samples with higher weights are more likely to be resampled, creating the third layer of circles in the figure.

Looking at the third layer of circles, it is clear that we may end up with multiple redundant samples after resampling. If we were to continuously weight and resample these particles, we would eventually converge on just one particle. The nearness of that sample to the actual peak of the posterior distribution is limited by the granularity with which we created our initial samples.

Not having enough samples to cover the highest probability regions is often referred to as *sampling depletion*. One method of combating this problem is to perturb

the resampled particles by a gaussian distribution with small variance. The process of gaussian resampling or *regularization* will be discussed further in section 4.2. The gaussian resampling and re-weighting of those samples according to their probability given the evidence completes the figure and the basic particle filtering algorithm. The algorithm may be repeated as many times as needed.

The particle filtering algorithm is well suited for estimation in a time-varying system. To apply particle filtering to such a system, one merely needs to propagate the samples after each time-step by the state transition equation. Since each sample is a complete instantiation of the state at that time-step, any formulation of system dynamics, whether it be linear or nonlinear, will accurately represent estimates of the future state. The new samples created by propagating the previous time-steps samples, can then be re-weighted by the evidence of the next time step and resampled etc.. The likelihood of the observation given the state is determined by the measurement equation.

From the algorithm, one can see there are no assumptions of linearity of either the state transition or measurement equations. The state represented by the samples may also be nongaussian and multi-modal. It is for these reasons that we are studying this technique for sequential inference as an alternative to extensions of Kalman Filtering. Although, we are no longer constrained by approximations of linearity or unimodality of the Kalman Filtering methods, we have now must consider approximations due to stochastic sampling. Chapter 4 will discuss these considerations and techniques for improving the efficiency of sampling with a finite number of particles.

Chapter 4

Particle Filtering Issues

With an infinite number of samples, the particle filtering algorithm theoretically approaches the optimal Bayesian estimate of the posterior probability distribution. However, practical implementations of this algorithm must limit the number of samples depending on how quickly the algorithm must perform inference. As a result, there are limitations created by the approximations of finite sampling.

The two main sampling considerations of particle filtering are referred to as *sample degeneracy* and *sampling depletion.*[1] Both are concerned with modeling the higher probability areas of the posterior distribution.

Sample degeneracy signifies the efficiency with which the samples model the desired distribution. If there are many particles with very low probability weights, those low-weight samples do not accurately model the high probability areas of the distribution. Usually the process of resampling will rid us of many low probability samples, since the low probability samples will be resampled infrequently.

Sample depletion, however, refers to the case when there are no samples in the high probability areas we wish to model. Once samples have been depleted from that area, it is very difficult to gain an accurate representation of the distribution. Due to the inaccuracy of sampling, a posterior distribution with sharp peaks will be difficult to model since the probability of choosing samples on those peaks, using a more uniform proposal distribution, is very low. Hence, for sharp posterior probability distributions too frequent resampling and sample depletion are major concerns.

Particle filtering techniques will have most difficulty when the likelihood of the newly observed evidence is low, since the likelihood that we have sampled the regions that could result in those observations will also be low. However, this deficiency is being traded for the efficiency of the many cases where the observations will be likely according to our prior beliefs. In order to improve our particle filter as a robust technique for inference, there are some methods to help combat sample depletion while attempting to keep sample degeneracy to a minimum.

4.1 Controlled Resampling

It is clear that sample depletion and degeneracy are closely linked to the frequency with which we resample our particles. If we resample too often, samples in less likely probability areas will be lost and we increase the possibility that future observations cannot be explained by our remaining samples. However, resampling too infrequently will result in many samples with negligible probability weights; particles which could better be used to sample higher probability regions of the distribution.

One idea for controlling the frequency of resampling is to link the frequency to the variance of the particle weights. An approximate measure of the number of effective particles, N_{eff}, where N is the number of particles and w_i represents the weight of particle i, is below.

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} w_i^2} \tag{4.1}$$

When N_{eff}, the approximate number of particles, falls below a chosen threshold, one should resample the particles.[1] The effect is to resample only when there are too many samples of very low weight and not in situations where the reasonably probable particles are likely to be depleted. However, choosing the threshold for resampling is not an easy task. This tuning parameter can have a major effect on the performance of the filter because of the serious consequences of sample depletion.

4.2 Regularization

Another method for combatting errors due to sampling is often referred to as regularization. This technique involves perturbing samples using a gaussian distribution after resampling. In the particle filtering algorithm shown in figure 3-2, this step is represented by the slight shifting of resampled particles in the fourth group of samples from the top.

The problem with particle filtering without regularization is that in systems where x_t is continuous, after each resampling event, our samples of x can only estimate x's value with precision limited by the locations of the original samples. Without gaussian perturbations, even if the samples narrow themselves down to one region with little uncertainty, the region cannot be a more precise estimate of x than any of the original samples.

These random perturbations are necessary in order to allow finer exploration of the continuous space. In situations where the number of samples used do not cover the state space with the precision desired, this technique is very useful in countering the natural inaccuracies due to under-sampling.

However, in highly peaked posterior densities, this technique will not help to find sharp peaks since samples must have weights large enough to be resampled in the first place before the regularization process can perturb them. As a result, particle filters will be more likely to converge on local minima in these rapidly changing, peaked distributions. In addition the gaussian perturbations should remain relatively small in order to allow eventual convergence to peaks and not shift samples too far away from the desired areas. Regularization, or gaussian resampling is a useful measure to increase the precision and resolution of a particle filter, but its ability to combat sample depletion is limited.

4.3 Weight Update Lower Bounds

Another technique to combat sample depletion is to place a limit on how small the weight update, $P(y_t|x_t)$, can be. For the gaussian additive noise we are concerned with in this study, this is similar to using a truncated gaussian to re-weight samples given the new evidence.

The results of placing a lower bound on new probability weights can help to combat the effects of outlier observations. A particularly noisy observation may cause sample depletion in the true posterior distribution if the weights of particles are allowed to fall too quickly. Generally this technique is not needed when there are many samples to cover the sample space. However, the costs of using the technique are low and convergence will generally not be significantly slowed.

This chapter has described four techniques that may be used to counter the effects of insufficient sampling and sampling depletion. There are many other techniques, including adaptive particle filtering [6] and auxiliary particle filtering [1]. Each technique is tailored to a different set of circumstances. My intention was to present techniques that were most relevant to the application studied in this thesis. Chapter 5 will discuss the Spinning Discs Problem for which these particle filtering algorithms were used. Chapter 6 will discuss the results of particle filtering for the Spinning Discs scenarios.

Chapter 5

The Spinning Discs Problem

The Spinning Discs problem is a test scenario, developed by David Cebula and Keh-Ping Dunn at the MIT Lincoln Laboratory, as an exampling filtering problem that can be described through a Dynamic Bayesian Network. A disc spins while an observer sitting at the edge of the disc sees the color intensity of the disc as each portion spins by. The goal is, given a discrete set of possible disc color patterns, to estimate which pattern we are observing, the rate at which the disc is spinning, and the current angle of the disc.

The rate at which the disc spins as well as the pattern of the disc remains static. There may be gaussian additive noise associated with each observation, but we know the variance of that noise.

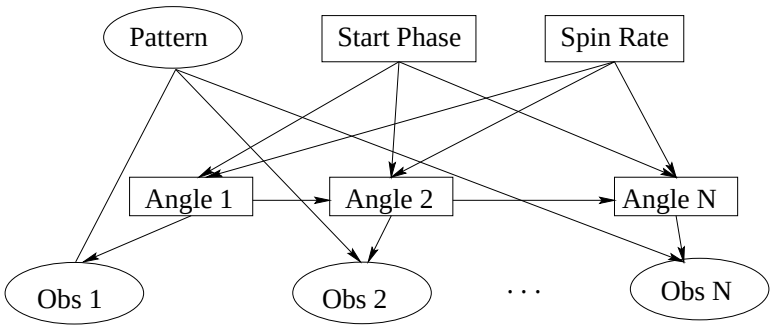


Figure 5-1: Dynamic Bayesian Network for the Spinning Discs Scenario

5.1 Transition and Measurement Model

Transition Equation:

$$\mathbf{x_t} = \begin{bmatrix} pattern \\ spin - rate \\ angle \end{bmatrix} = A\mathbf{x_{t-\delta t}} + \mathbf{n_t}$$

Observation Equation:

$$y_t = \begin{bmatrix} color \end{bmatrix} = h(\mathbf{x_t}) + v_t$$

The state dynamics and measurement models are given above in the transition and observation equations. The state is a linear function of the previous state plus gaussian process noise, n, which for spinning discs is taken to be zero. In spinning discs, the linear transformation, in the matrix A, represents no change in pattern and spin-rate, and an increment in angle equal to the spin-rate multiplied by the change in time.

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \delta t & 1 \end{bmatrix}$$

The system dynamics is clearly linear and has no theoretical noise, since we have taken $\mathbf{n_t}$ to be zero. However, the measurement model is nonlinear and has the additive gaussian noise, v. The function, $h(\mathbf{x_t})$, depends on the pattern and the angle contained in the state variable x. Since the patterns are not necessarily linear with angle, this equation can be nonlinear. As a result, the densities implied by certain color observations can be multi-modal, motivating our need for Particle Filtering.

5.2 Possible Patterns

First the particle filter was compared to a Kalman filter with the linear measurement model shown in figure 5-2. Then nonlinear patterns were used to test the particle filter in the nonlinear spinning discs scenario. The cosine pattern in figure 5-3 was used to test estimation alone. In addition, two highly peaked patterns were used for estimation and discrimination. These patterns are shown below in figure 5-4 and 5-5.

The results of the particle filter on each of these patterns are discussed in Chapter 6, Results.

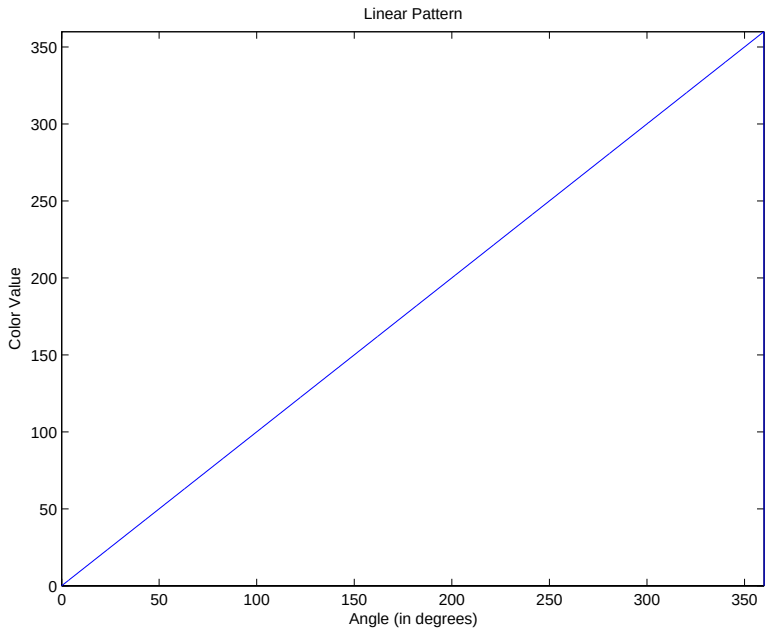


Figure 5-2: Linear Pattern (Linear Measurement Model) for Spinning Discs Estimation

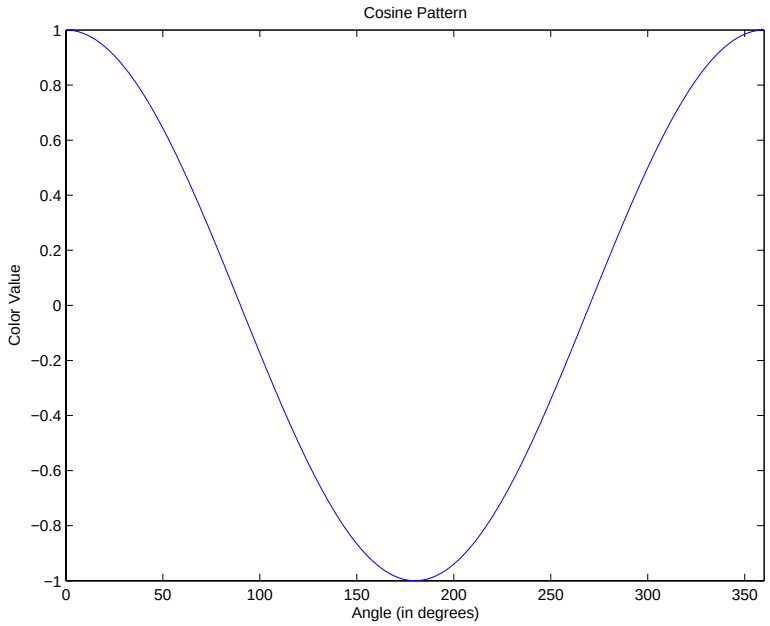


Figure 5-3: Cosine Pattern (Nonlinear Measurement Model) for Spinning Discs Estimation

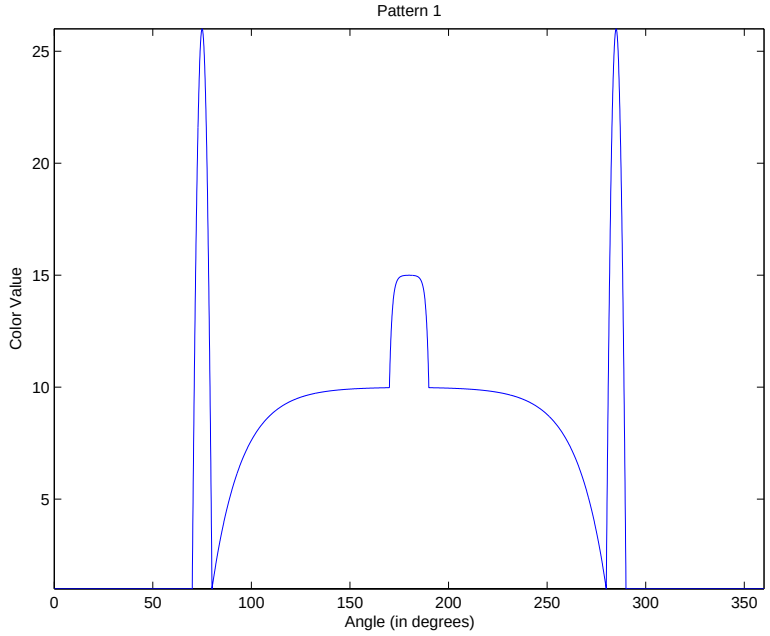


Figure 5-4: Nonlinear Pattern 1 for Spinning Discs Discrimination and Estimation

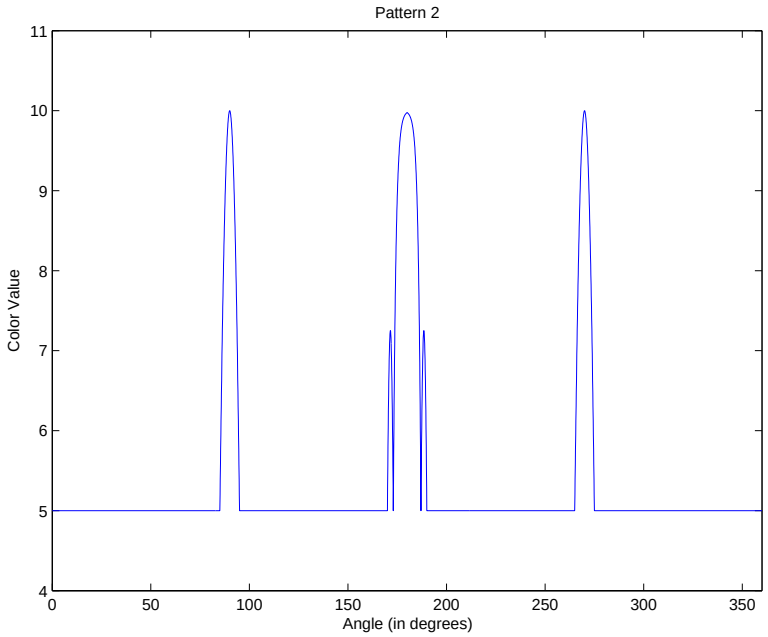


Figure 5-5: Nonlinear Pattern 2 for Spinning Discs Discrimination and Estimation

Chapter 6

Results

This chapter presents the results of tests of the particle filter on the spinning discs system. The first section, 6.1, compares the 3000-sample particle filter to the Kalman filter for spinning discs with linear patterns. Since the Kalman filter is already the least squares solution, the first section is a validation of the ability of the particle filter to deal with linear filtering and estimation problems. It also provides a clear comparison of the performance of the two filters that demonstrates the sensitivity of the particle filter to its various parameters and the possible limitations of particle filtering.

Section 6.2 tests the performance of the particle filter on the nonlinear spinning discs problem. The cosine pattern is used, creating a nonlinear measurement model and multi-modal posterior probability distributions. In order to compare the performance of various particle filters, the KL divergence of each particle distribution is calculated with a very finely sampled grid-based filter. This section examines the effects of the number of particles, resampling threshold, regularization variances, measurement noise level, and number of time-steps on the performance of the particle filter.

Finally section 6.3 evaluates the ability of the particle filter to discriminate among more highly peaked nonlinear patterns. The estimation of angle and spin-rate will also be examined in these challenging patterns.

6.1 Comparison to Kalman Filter

In order to validate the design of the particle filter, it was first compared with a Kalman filter for the linear filtering problem. The spinning disc pattern used for these comparisons was a linear function of angle, as shown in figure 5-2. In such cases, where the measurement and transition models are linear, the Kalman filter is the unbiased, efficient estimator and least squares optimal solution.[11]

The spinning discs Kalman and particle filters simultaneously received the same noisy observations from which to estimate the spin-rate and angle. The system parameters and particle filter parameters are described at the beginning of each comparison followed by graphs of the resulting estimates and errors of each filter.

6.1.1 Kalman Filter vs. 3000-Sample Particle Filter

Table 6.1: Parameters of the System

Pattern	Linear
Number of Time-steps	100
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	1/4
Angle Process Noise	0
Spin-rate Process Noise	0

Table 6.2: Parameters of the Kalman Filter

Angle Process Noise	1/10
Spin-rate Process Noise	1/100

There is no process noise in the linear spinning discs system. However, Kalman filters will often have non-zero terms for process noise even if the there is none in the underlying system. In these cases, the process noise parameters used in the Kalman filter are to enhance performance. By adding a little process noise, the Kalman filter

Table 6.3: Parameters of the Particle Filter

Number of Particles	3000
Resampling Threshold	1/4
Angle Regularization Variance	1/2
Spin-rate Regularization Variance	1/4

may converge more quickly to the correct estimate. The very small process noise used in the Kalman filter for these tests is listed in Table 6.2.

Particle filtering parameters, such as the number of samples, resampling threshold, and the variances of the regularization kernels, can significantly affect the performance of the filter. Since sample depletion is less likely to occur when there are many particles, the performance of the particle filter was first compared to the Kalman filter using 3000 particles. The particle filter's parameters are outlined in Table 6.3.

The reasoning for the use of a resampling threshold is explained in section 4.1, where the relationship between resampling and sample depletion is discussed. Here, the resampling threshold represents a percentage of the total number of particles, N_{eff}/N. When the calculated N_{eff}/N is less than this pre-set higher bound, the particle filter will resample.

The technique of regularization is also discussed in section 4.2. Lower resampling thresholds and higher regularization variances are intended to combat sampling depletion and sampling difficulties. However, the main factor in preventing sample depletion is the number of particles used, which has been chosen to be large for this comparison of performance of the particle filter with the Kalman filter.

Figure 6-1 shows the comparison of the particle filter and Kalman filter's estimates of angle with time. Although the estimates differed initially, they quickly converged. This shows that, with 3000 particles and enough observations, the particle filter can produce a very close to optimal estimate of angle for this linear filtering problem. A comparison of the covariances of the particle and Kalman filter's estimates of angle is shown in figure 6-6.

The particle filter also did well on estimating the spin-rate of the disc, as shown

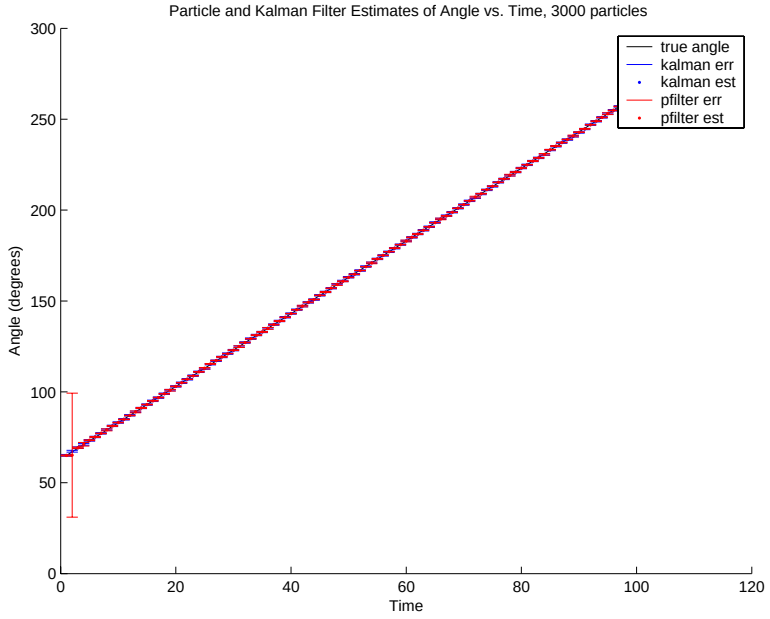


Figure 6-1: Comparison of the Angle Estimates of the Kalman and Particle Filter with 3000 samples, with angle and spin-rate regularization variances of 1/2 and 1/4, respectively. The error bars represent the standard deviation of each estimate.

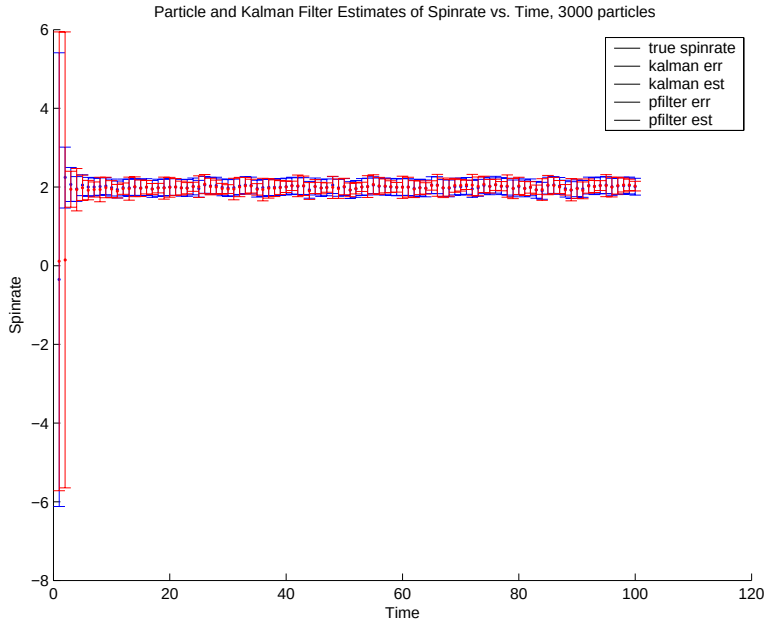


Figure 6-2: Comparison of the Spin-rate Estimate vs. Time of the Kalman Filter and 3000-sample Particle Filter with angle and spin-rate regularization variances of 1/2 and 1/4, respectively. The error bars represent the standard deviation of each estimate.

in figure 6-2. This figure also shows the error bars of the estimates, which are the standard deviations of the estimate of the spin-rate. Initially the covariance of the particle filter estimate is much greater than the Kalman filter's spin-rate covariance. After a few time-steps, the spin-rate estimate and its covariance converge closely to that of the Kalman filter.

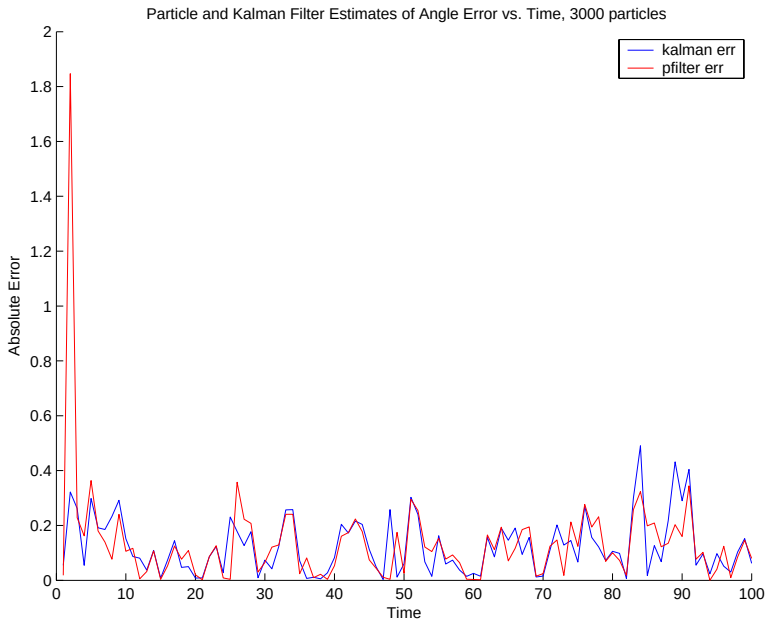


Figure 6-3: Comparison of the Absolute Error of The Angle Estimate vs. Time for the Kalman Filter and 3000-sample Particle Filter, with angle and spin-rate regularization variances of 1/2 and 1/4, respectively.

In order to see how closely the particle filter matched the Kalman filter's estimate of angle, figure 6-3 shows the absolute error, $|\hat{\mathbf{x}} - \mathbf{x}|$, of the particle filter versus the absolute error of the Kalman filter's angle estimate. After approximately 8 time-steps, the error of the particle filter follows the error of the Kalman filter quite closely, never differing by more than half of a degree.

The error of a particle filter's mean angle estimate is strongly related to the regularization variances. Figure 6-4 shows the error for a particle filter with the same parameters, but with the higher regularization variances of 1 and 1/2. The result is greater error, since the regularization adds uncertainty to the particle distribution. For the particle filter with higher regularization parameters, the time-steps in which

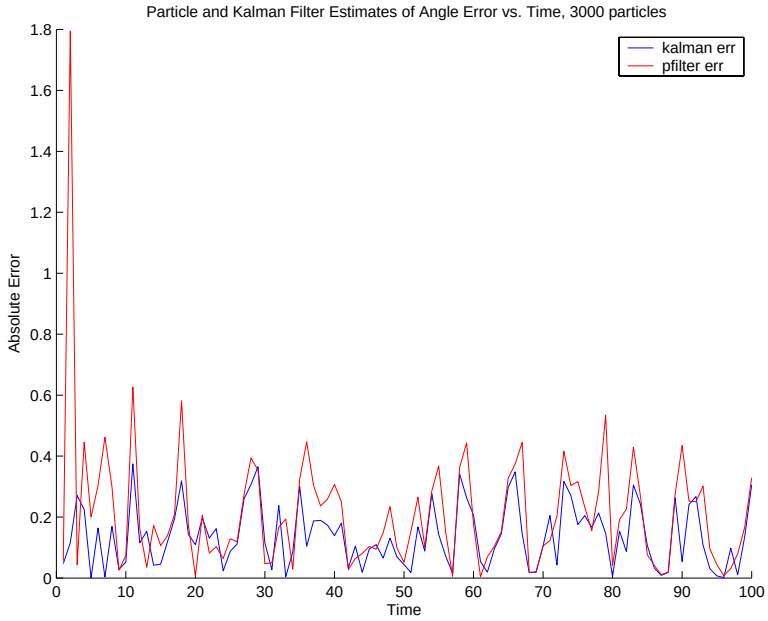


Figure 6-4: Comparison of the Covariance of The Angle Estimate vs. Time for the Kalman Filter and 3000-sample Particle Filter, with angle and spin-rate regularization variances of 1 and 1/2, respectively.

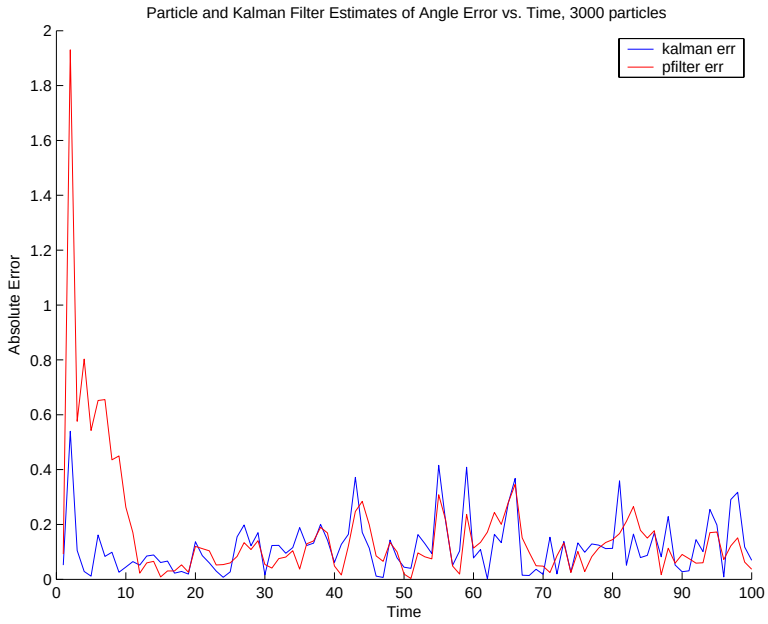


Figure 6-5: Comparison of the Covariance of The Angle Estimate vs. Time for the Kalman Filter and 3000-sample Particle Filter, with angle and spin-rate regularization variances of 1/4 and 1/10, respectively.

the Kalman filter's error increases are the same as the time-steps in which the additive noise value, v_t, is greater. During those times, the particle filter estimate deviates from the true value even more than the Kalman filter estimate, showing greater sensitivity of the particle filter to noisy observations. The reason for this is regularization.

The regularization kernels, used to smooth the particles upon resampling, add uncertainty to the estimate. This uncertainty, represented by an artificially large spread of samples around the mean, results in the existence of samples further away from the mean. During a noisy observation, these deviant samples receive higher probability weights than those near the mean. The overall result is to allow more deviation of the mean estimate with noisy observations.

In order to confirm that these large deviations in angle estimate were, in fact, due to the the angle and spin-rate regularization kernels, the experiment was re-run with lower regularization variances and otherwise identical parameters . The lowered variances did indeed produce less error, as shown in figure 6-5. However, lowering variances too much may cause the particle filter to become overconfident. Without regularization the particle distribution will generally have a smaller covariance than the true distribution. As a result, the particles may diverge from the correct estimate, increasing the estimate error. This is the scenario referred to as sampling depletion.

A final comparison of the 3000-sample particle filter to the Kalman filter is to examine the covariance of the angle estimates over time. Figure 6-6 shows the covariance of the particle filter and Kalman filter estimates for the regularization variance of 1/2 and 1/4. After about 10 time-steps, the covariances of the two filters became very similar. Although the particle distribution had a much higher covariance than the posterior distribution in the first few time-steps, because of the inaccurate proposal distribution, it quickly converged to covariances that did not differ from the Kalman filter's covariance by more than 1/2 of a degree.

In addition, we can discern a somewhat regular pattern for the difference. The periodically increasing covariance is related to the frequency of resampling. After the particles are resampled the covariance of the estimate is increased due to the added uncertainty of the angle and spin-rate regularization variances. However, the

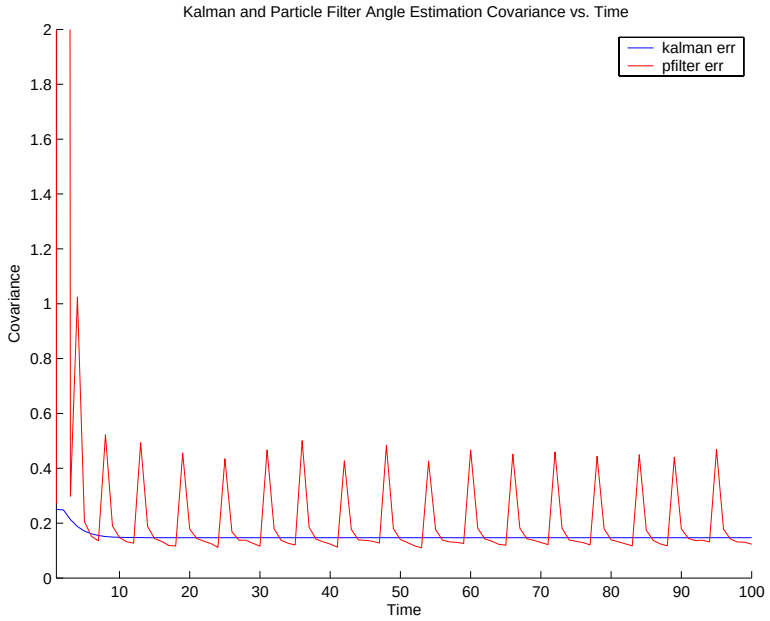


Figure 6-6: Comparison of the Covariance of The Angle Estimate vs. Time for the Kalman Filter and 3000-sample Particle Filter, with angle and spin-rate regularization variances of 1/2 and 1/4, respectively

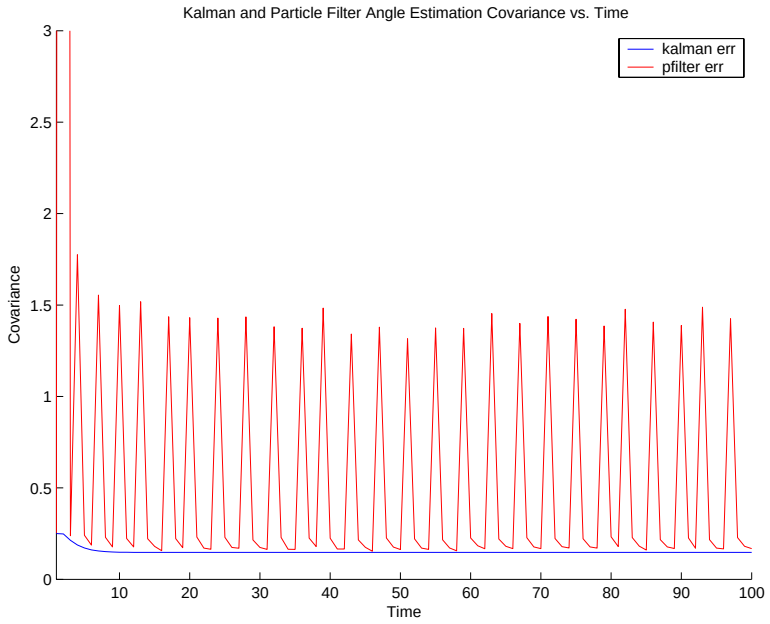


Figure 6-7: Comparison of the Covariance of The Angle Estimate vs. Time for the Kalman Filter and 3000-sample Particle Filter, with angle and spin-rate regularization variances of 1 and 1/2, respectively

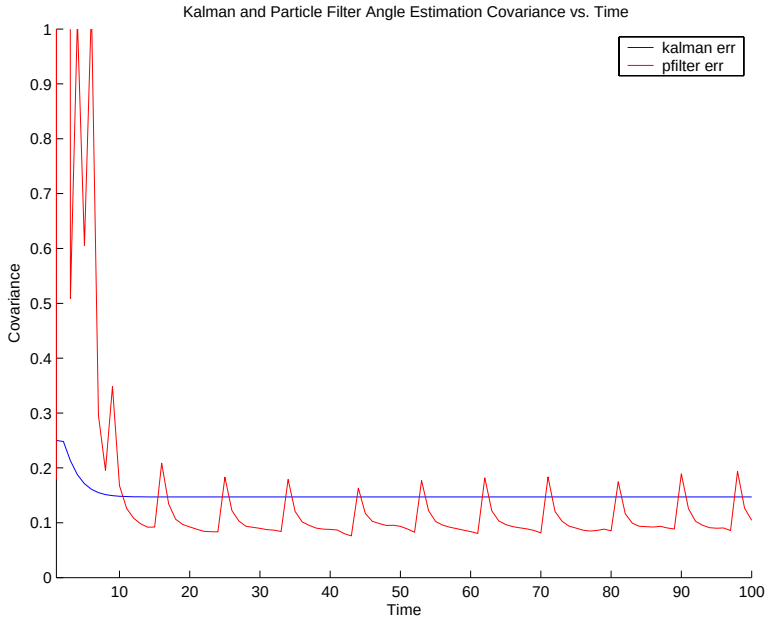


Figure 6-8: Comparison of the Covariance of The Angle Estimate vs. Time for the Kalman Filter and 3000-sample Particle Filter, with angle and spin-rate regularization variances of 1/4 and 1/10, respectively

covariance decreases during time-steps when the particles are not resampled as the samples are re-weighted by the observations.

The comparison of the covariances of the particle and Kalman filters may explain the differences in the angle estimation errors as it depends on the regularization variance. Figure 6-7 shows the covariance of the particle filter with regularization variances of 1 and 1/2 compared with the covariance of the Kalman filter estimate. Here the covariance is always greater than that of the Kalman filter, which may prevent sample depletion. This also helps explain the increase in angle estimate error with noisy observations shown in figure 6-4. The increased covariance makes the particle filter more likely to change its estimate with noisy observations.

Figure 6-8 shows the covariance corresponding to the particle filter with lower regularization variances. Clearly, the difference between the covariances is greatly reduced from the case in which the smoothing parameters were higher. However, this time we can see that after not resampling for a few time-steps the covariance of the particle filter's estimate is too low. This artificially low covariance is an inaccuracy

inherent to discrete sampling and shows why additional uncertainty, in the form of regularization, is needed. The particle filter with a covariance lower than that of the Kalman filter is over confident about its estimate. As a result, it will pay less attention to observations than it should, possibly leading to divergence.

The results of comparing the 3000-sample particle filter to the Kalman filter demonstrate the particle filter's ability to approximate the Kalman filter on linear problems. However, the particle filter is intended for use in non-linear problems which may have multi-modal, non-gaussian posterior distributions. Despite this contradiction, the comparison of the particle filter to the Kalman filter for linear problems is still useful. The linear particle filter can illustrate the various sensitivities of this type of filter to its many parameters, while providing an easy comparison to the optimal solution achieved by the Kalman filter.

The next two sections will examine two of the main factors that affect the performance of the any particle filter. Section 6.1.2 will study the performance of the particle filter on linear spinning discs with respect to the number of samples. Then section 6.1.3 compares the particle and Kalman filters in linear systems with varying levels of noise. Although these results help to point out possible weaknesses of the particle filter, they do not directly apply to nonlinear problems. Therefore, section 6.2 will also examine these issues for spinning discs with the nonlinear, cosine measurement model.

6.1.2 Kalman vs. Particle Filter for Various Numbers of Samples

The comparison of the 3000-sample particle filter to the Kalman filter revealed a very close match. After enough time, with a non-trivial measurement noise level, and with a large number of samples, the particle filter performs very well. However, more difficult, nonlinear problems, often require more samples to accurately model the posterior distribution. It is usually too computationally expensive to use sufficiently

many particles as to make sampling issues so trivial. This section will investigate the performance of particle filters with fewer samples compared to the Kalman filter.

Table 6.4: Parameters of the Particle Filters

Number of Particles	100,500,1000,3000
Resampling Threshold	1/4
Angle Regularization Variance	1
Spin-rate Regularization Variance	1/2

Table 6.4 details the filter parameter values for these tests of performance versus the number of samples. The parameters are the same as those in the previous section that used 3000 particles.

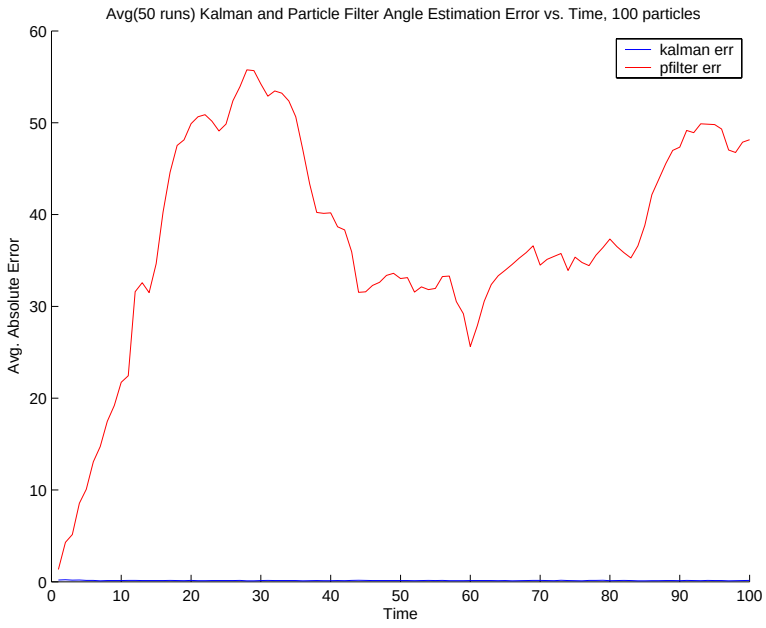


Figure 6-9: Comparison of the Average Angle Estimate Error vs. Time for the Kalman Filter and Particle Filter with 100 particles, averaged over 50 runs

Figures 6-9, 6-10, 6-11, and 6-12 show the average absolute error of the particle filter's estimate of angle with 100, 500, 1000, and 3000 samples, respectively. The number of runs used to calculate the averages was greater for the filters with less particles, due to the greater amount of time needed to run particle filters with more samples.

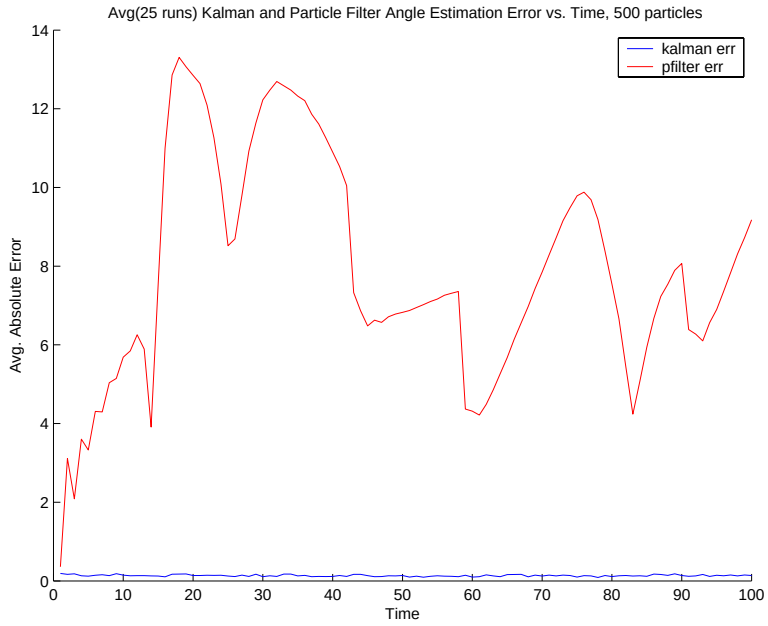


Figure 6-10: Comparison of the Average Angle Estimate Error vs. Time for the Kalman Filter and Particle Filter with 500 particles, averaged over 25 runs

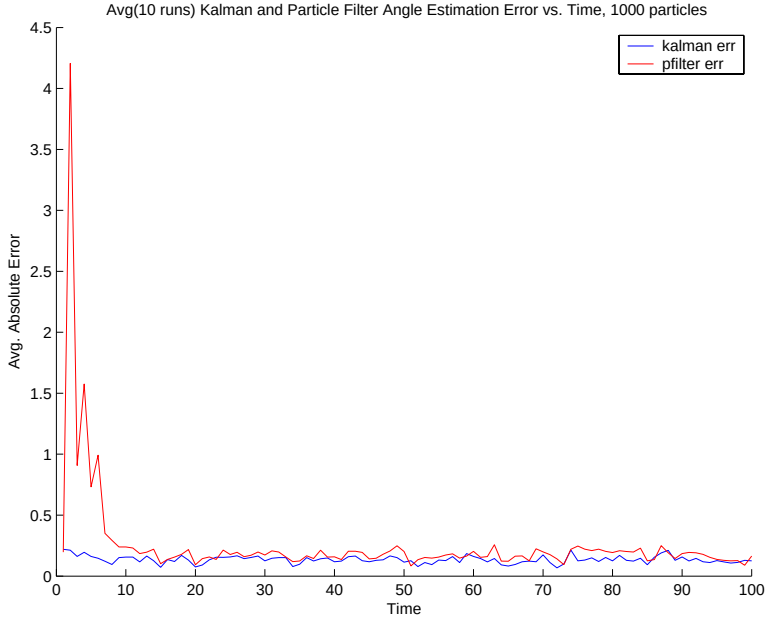


Figure 6-11: Comparison of the Average Angle Estimate Error vs. Time for the Kalman Filter and Particle Filter with 1000 particles, averaged over 10 runs

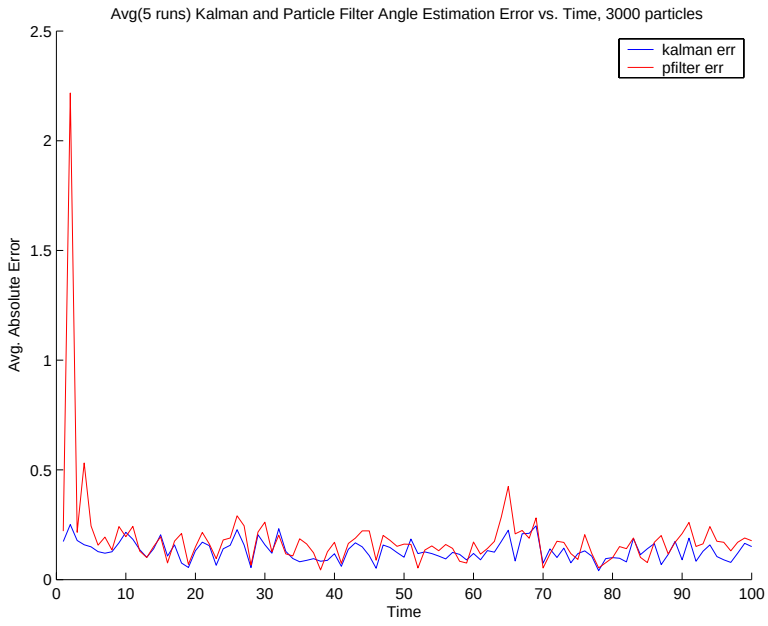


Figure 6-12: Comparison of the Average Angle Estimate Error vs. Time for the Kalman Filter and Particle Filter with 3000 particles, averaged over 5 runs

These results clearly support the prediction that the absolute error will decrease as the number of particles increase. They also demonstrate the consequences of not having enough samples. The graphs of error with 100 and 500 particles, figure 6-9 and figure 6-10, are good examples of sample depletion. Without enough particles, the high probability regions are often completely missed by the initial samples. Such severe sample depletion makes recovery of the high probability regions unlikely since the amount the particles can spread, and search for new answers, is limited by the regularization variances.

The results of plotting the absolute error versus the number of samples highlight the importance of having enough particles to avoid severe sample depletion. When severe sample depletion occurs, the particle filter usually cannot recover the correct estimate. In addition, sample depletion occurs more often in the particle filters with less particles. These two factors combine to drastically increase the average error of filters with insufficient samples since they are more prone to sample depletion.

6.1.3 Kalman vs. Particle Filter in Various Levels of Noise

Another consideration affecting the performance of the particle filter is the level of noise inherent in the system. Except in cases where the linear model doesn't fit the true model, the Kalman filter always does better in lower measurement noise. In contrast, particle filters sometimes performance worse in situations where there is less noise. These tests show that once the measurement noise level is small enough, the spinning discs problem becomes sufficiently degenerate to be make modeling the posterior distribution with samples very difficult.

In order to compare the performance of particle filters to the Kalman filter in various levels of measurement noise, a 1000-sample particle filter was tested in systems with measurement noise levels of 1/4, 1/10, 1/25, 1/50, and 1/100. The complete system and particle filter parameters are listed in tables 6.5 and 6.6.

Table 6.5: Parameters of the System

Pattern	Linear
Number of Time-steps	100
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	1,1/4,1/10,1/25,1/50,1/100
Angle Process Noise	1/100
Spin-rate Process Noise	1/200

Table 6.6: Parameters of the Particle Filter

Number of Particles	1000
Resampling Threshold	1/4
Angle Regularization Variance	1
Spin-rate Regularization Variance	1/2

Comparing the trial with measurement noise of 1, shown in figure 6-13, with the trial with measurement noise of 1/4, figure 6-14, we clearly see both the Kalman and particle filter improve with less measurement noise. The trend of decreasing average

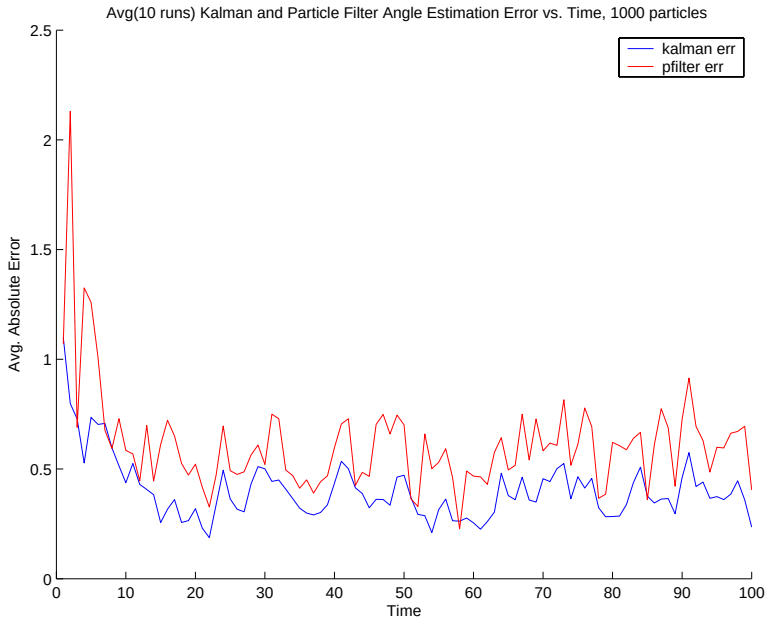


Figure 6-13: Comparison of the Average Angle Estimate Error for measurement Noise of 1. The Kalman Filter versus 1000-sample Particle Filter, averaged over 10 runs.

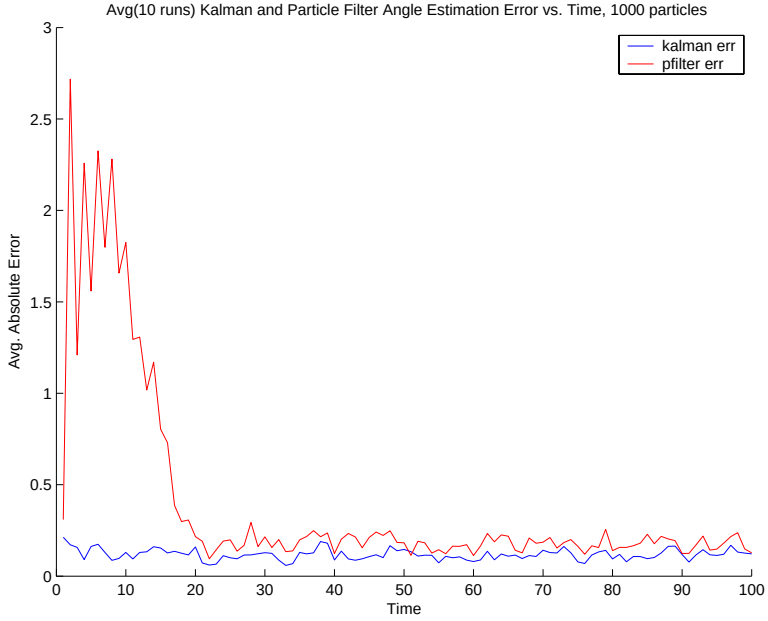


Figure 6-14: Comparison of the Average Angle Estimate Error for measurement Noise of 1/4. The Kalman Filter versus 1000-sample Particle Filter, averaged over 10 runs.

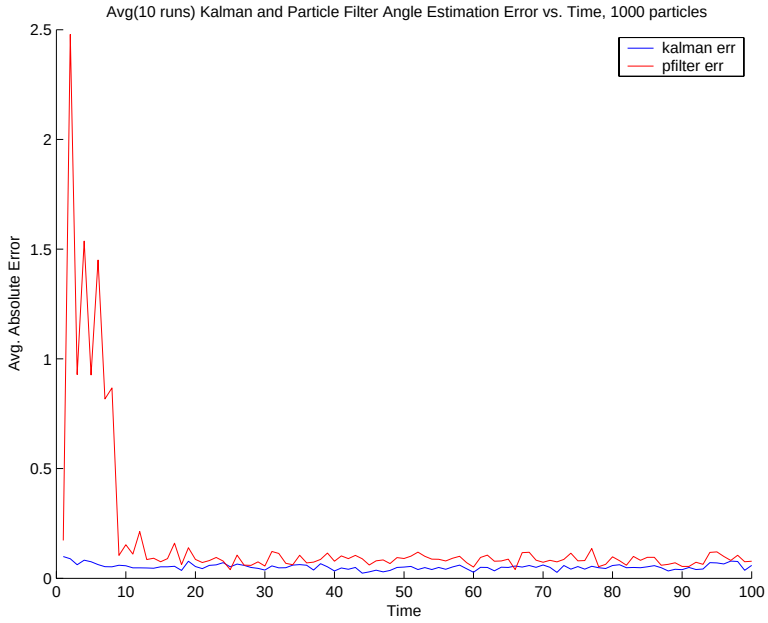


Figure 6-15: Comparison of the Average Angle Estimate Error for measurement Noise of 1/10. The Kalman Filter versus 1000-sample Particle Filter, averaged over 10 runs.

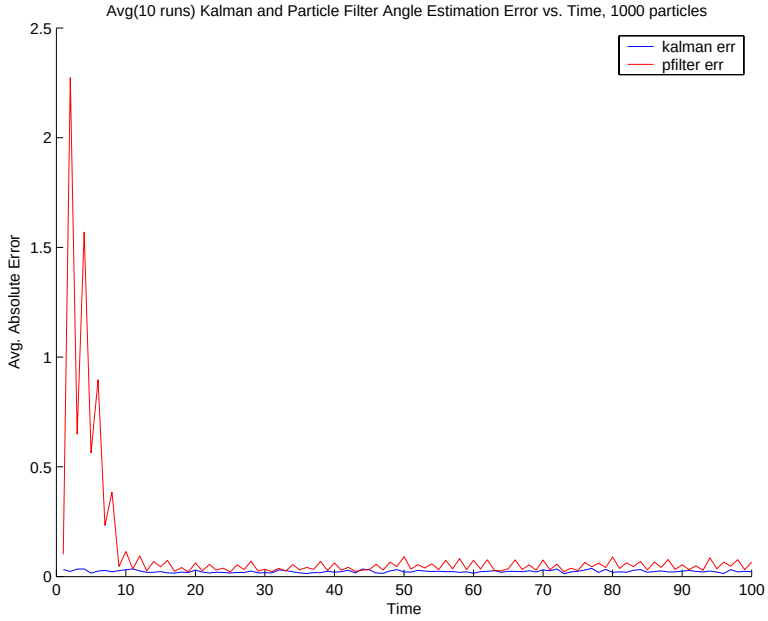


Figure 6-16: Comparison of the Average Angle Estimate Error for measurement Noise of 1/25. The Kalman Filter versus 1000-sample Particle Filter, averaged over 10 runs.

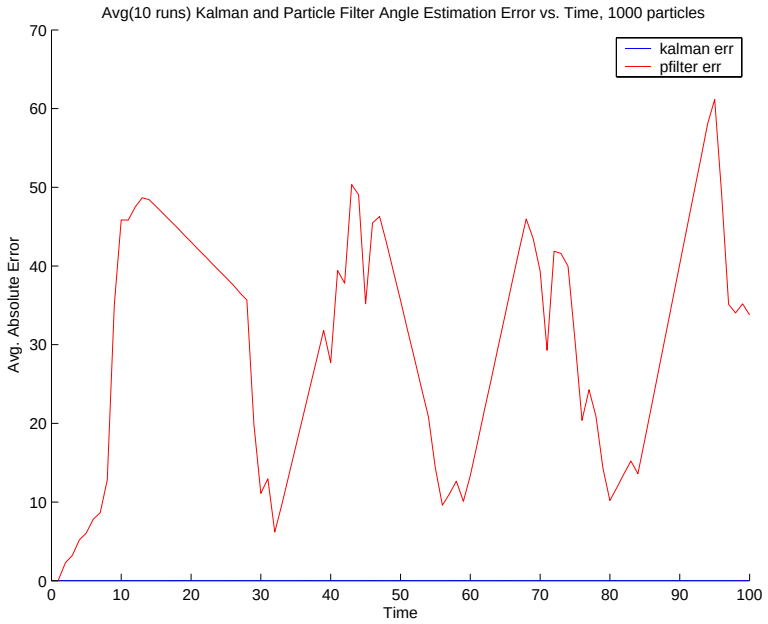


Figure 6-17: Comparison of the Average Angle Estimate Error for measurement Noise of 1/50. The Kalman Filter versus 1000-sample Particle Filter, averaged over 10 runs.

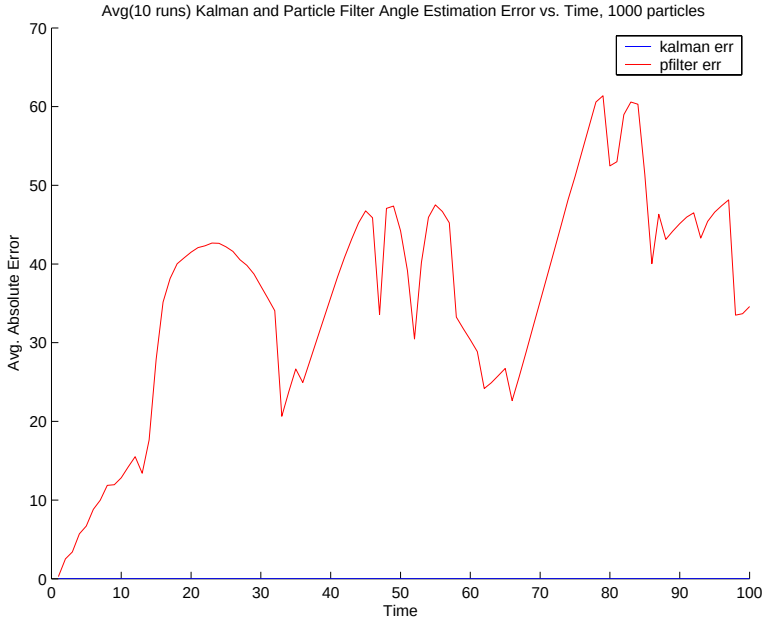


Figure 6-18: Comparison of the Average Angle Estimate Error for measurement Noise of 1/100. The Kalman Filter versus 1000-sample Particle Filter, averaged over 10 runs.

absolute error with decreasing measurement noise continues for the even lower noise level of 1/10 and 1/25 displayed in figures 6-15 and 6-16.

However, once the noise level is as low as 1/50, there is a drastic increase in absolute error for the 1000-sample particle filter. Figure 6-17 shows an average absolute error orders of magnitude greater than that for the noisier trials, when the measurement noise is only 1/50. In addition, this dramatic increase in error persists as the noise level drops even further to 1/100 in figure 6-18.

While the Kalman filter has less absolute error as the measurement noise decreases, the particle filter does not display the same behavior. The reason for this is inadequate sampling. As the noise level drops, the posterior distribution of spin-rate and angle given observations of color becomes sharper. The sharper distribution is merely a result of there being less uncertainty since the measurements are less noisy.

As a result, the 1000 samples of the particle filter, initially sampled with a uniform proposal distribution, are less likely to fall in the correct region of the state space. Sample depletion occurs because the prior distribution modeled by the particle filter cannot explain the posterior distribution due to a new observation. When the noise level is very low, the posterior distribution, after only a couple observations, is much sharper than the uniform proposal distribution modeled by the particles. After sample depletion has occurred, the correct state is extremely difficult to recover, making this an unrecoverable error which will increase the error of the particle filter by orders of magnitude.

The results may seem obvious. In problems where there is little uncertainty, guessing according to a bad prior is clearly going to be a very wasteful strategy. This is exactly why so much of the general particle filtering literature emphasizes using the best proposal distributions possible. In very low noise cases, instead of wasting most of the 1000 samples by using a uniform prior, a better strategy might be to see the first color observation and then sample a distribution around that value.

However, for many problems for which particle filters are used, we do not know the shape of the posterior distribution ahead of time, so choosing a good proposal distribution can be problematic. A constant worry of particle filtering is that if we

receive too much information at once, the inaccuracies of sampling will most likely show themselves in enormous errors.

6.2 Cosine Pattern

Measuring the performance of the particle filter for nonlinear patterns is not as simple as comparing its results to that of the Kalman filter. In the nonlinear case, the Kalman filter is not an optimal[1] solution for finding the maximum likelihood estimate or full posterior density. For the nonlinear spinning discs problem, we used compared the results of the particle filter to a grid-based filter with very fine samples.

The goal was to gain a model of the posterior distribution that is more accurate than the distribution of the particle filter. A finely sampled grid-based filter provided this more accurate model of the posterior density of angle and spin-rate. Each sample in the grid of the final angle and spin-rate was reweighted by its probability given all the observations, which generated a probability mass function of the posterior distribution.

In order to make a grid of samples that can be used with observations at different times, we relied on the ability to represent the angles and spin-rates for all times with the angle and spin-rate at a single time. This property is the consequence of having a deterministic transition equation. Without any process noise, the state at time t is a known linear transformation of the state at time $t-1$. Since each observation is independent, conditioned on the state, the probability weight for each sampled state in the grid-based filter reduces as is shown in equation 6.1. With a fine enough grid, this type of filter will very closely approximate the true posterior probability mass function.

$$p(\mathbf{x}|y_{1:n}) = \alpha \prod_{i=1}^{n} p(y_i|\mathbf{x}) \tag{6.1}$$

Now that we have a model of the posterior distribution, which we expect to be more complete and more accurate than the samples of the particle filter, we can

[1] Again optimal is being used in the Minimum Mean Square Error sense.

examine the Kullback-Leibler(KL) Divergence between the posterior distribution and a kernelized particle distribution. KL Divergence, also referred to as the relative entropy between two distributions, measures the average number of bits that are wasted by encoding events from a probably distribution, $p(x)$, with a code based on a less accurate distribution, $q(x)$.[3] The KL Divergence between $p(x)$ and $q(x)$ is given by equation 6.2 below. The approximate equality refers to the sum over samples of the possible states, which converges to the KL Divergence as the number of samples approaches infinity.

$$D(p(x)||q(x)) = \int p(x) log\left(\frac{p(x)}{q(x)}\right) dx \approx \sum_i p(x_i) log\left(\frac{p(x_i)}{q(x_i)}\right) \tag{6.2}$$

Next, the grid samples of the posterior probability *mass* function are converted to samples of the posterior probability *density* function. This is accomplished by dividing each of the sample probability weights by the area of each rectangle of the sample grid. The corrected weights will then correspond to samples of the probability density of the posterior.

Then the particles of the particle filter are converted to a Parzen density estimate by assuming each particle represents a Gaussian distribution with a constant kernel size. The whole distribution is the weighted sum of these Gaussian distributions. The leave-one-out maximum likelihood kernel size is found with equations 6.3 and 6.4 below. Each x_i represents one of the M particles and αp_j is the probability weight of particle x_j normalized to omit p_i. $N(a, b)$ refers to the zero-mean normal function evaluated at a, with variance b.

$$\sigma^2_{ML} = \overset{argmax}{\sigma^2} \left[\sum_{i=1}^{M} log\{\hat{p}(x_i; \sigma^2)\}\right] \tag{6.3}$$

where

$$\hat{p}(x_i; \sigma^2) = \sum_{j \neq i} \alpha p_j N(x_j - x_i; \sigma^2) \tag{6.4}$$

In order to quickly find the maximum likelihood kernel size, given by equation 6.3, a Golden Section Search was performed. This algorithm is a fast way of computing

the value which produces a local maxima of a continuous function bounded by a minimum and maximum value.[14] The minimum value for kernel size is equal to the minimum distance between two particles while the maximum kernel size is the maximum distance between any two particles. With the maximum likelihood kernel size, we have the posterior probability distribution of the particles, which we can then compare to the posterior distribution we obtained with the grid-based filter.

In order to compare the posterior distribution to the kernelized particle distribution, we calculate the KL Divergence between the posterior and the proposed distribution. The grid-sampled posterior distribution is $p(x)$ and a kernelized version of the particle distribution is $q(x)$. The KL Divergence in the following sections will refer to the comparison of a grid-sampled posterior with the weighted particle filter samples each evaluated with the maximum likelihood kernel size.

In the estimation of filter performance we relied on three assumptions. The first is that the samples of the grid-based filter are fine enough to model the posterior distribution more accurately than the samples of the particle filter. Secondly, we assume the calculation of KL Divergence from a finite number of samples will approach the true KL Divergence of the two distributions. The third assumption is that the Parzen density estimate of the particle distribution using the maximum likelihood symmetric gaussian kernel will be an accurate interpretation of the samples.

6.2.1 KL Divergence vs. Number of Samples

Table 6.7: Parameters of the System

Pattern	Cosine
Number of Time-steps	100
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	1/4
Angle Process Noise	0
Spin-rate Process Noise	0

Table 6.8: Parameters of the Particle Filter

Number of Particles	10,100,200,300,400,500, 1000,1500,2000,2500,3000
Resampling Threshold	5/6
Angle Regularization Variance	1/2
Spin-rate Regularization Variance	1/10

The goal of this set of tests was to characterize the performance of the particle filter in the nonlinear, cosine-patterned, spinning disc system as the number of samples used is changed. The noise parameters as well as the true spin-rate and initial angle are summarized in table 6.7. Except for the number of particles, the particle filtering parameters were held constant at the values given in table 6.8.

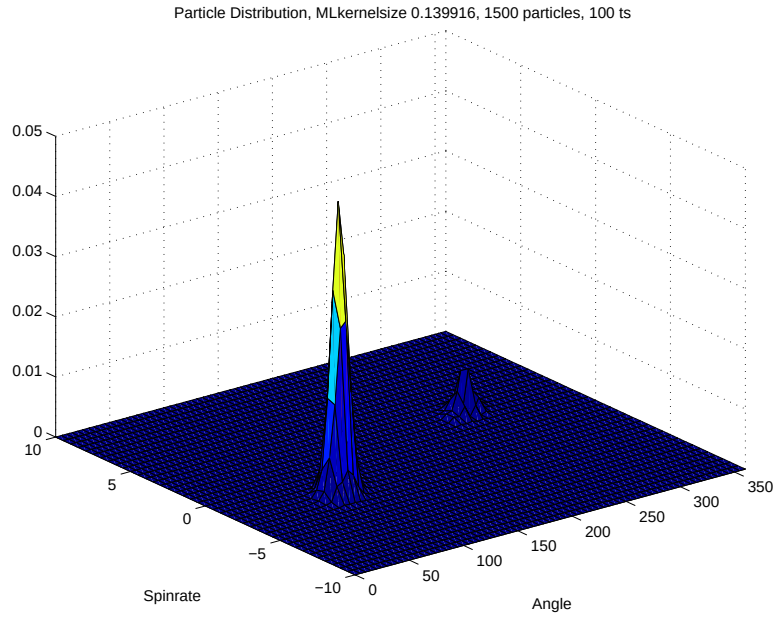


Figure 6-19: The Kernelized Particle Distribution for a Spinning Discs Particle Filter with 1500 particles and 100 time-steps

Figure 6-19 shows the kernelized sample distribution of the spinning discs particle filter using 1500 particles. We can see some sample depletion in this distribution, since the correct distribution should have two peaks at the +2 and -2 spin-rates.

Figure 6-20 graphs the grid-sampled posterior distribution for the same, 1500-particle test. Using all the 100 time-steps of observations, the posterior distribution

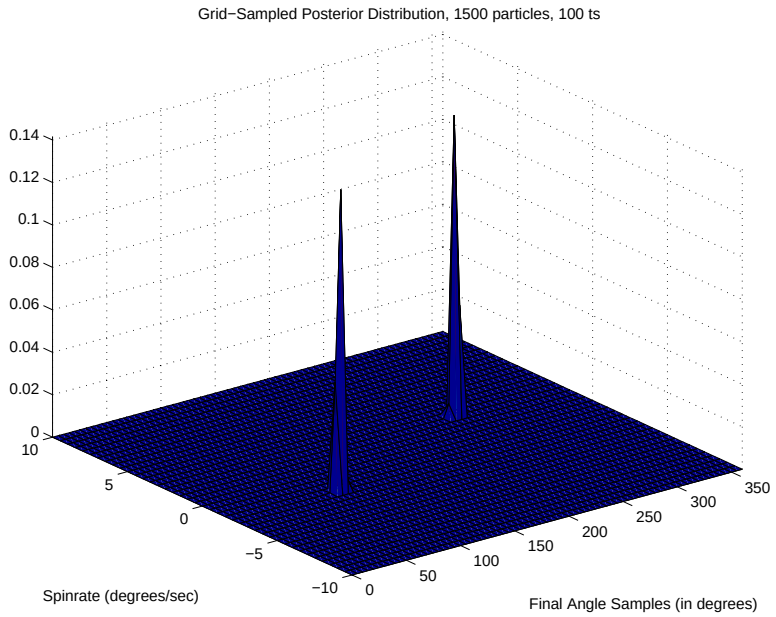


Figure 6-20: The Grid-Sampled Posterior Distribution, using the same 100 time-steps of observations as the 1500-Sample Spinning Discs Particle Filter

should be the two sharp peaks shown in this graph. The KL Divergence was calculated between the grid-sampled posterior and the kernelized particle distribution for 10 runs of each n-sample particle filter.

Figure 6-21 shows the 10-run average KL divergence between the grid-sampled posterior and kernelized particle filter distributions. We would expect the relative entropy to fall with the number of samples since there should be less sampling inefficiency when the filter has more particles, increasing accuracy of the sample distribution. However, the average KL Divergence shown in figure 6-21 does not strictly decrease with increasing numbers of particles. The unexpected increases can probably be explained by the large variances of these KL Divergences.

Figure 6-22 graphs the variance of the KL Divergence as it depends on the number of particles, for 10 runs. The variance is low for 10-sample particle filters, since these filter always have a close to maximum divergence. Then the variance is generally larger for the rest of the tests with fewer than 2000 particles since the chance of sample depletion is higher when there are less particles. When there are more than 2000 samples, the chances of sample depletion of effectively miniscule, so the variance

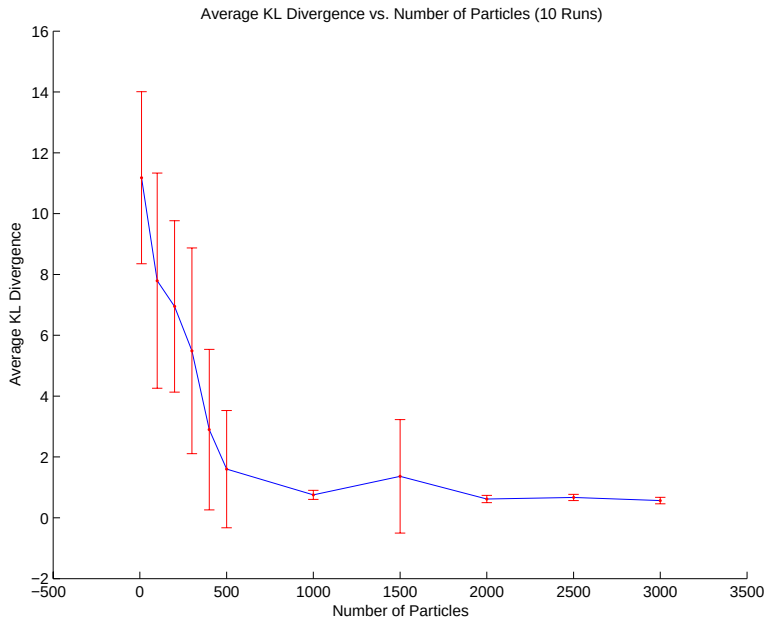


Figure 6-21: Average, over 10 runs, of KL Divergence of the Grid-Sampled Posterior Distribution and the Kernelized Particle Filter Samples. The red error bars represent the standard deviations of the averages.

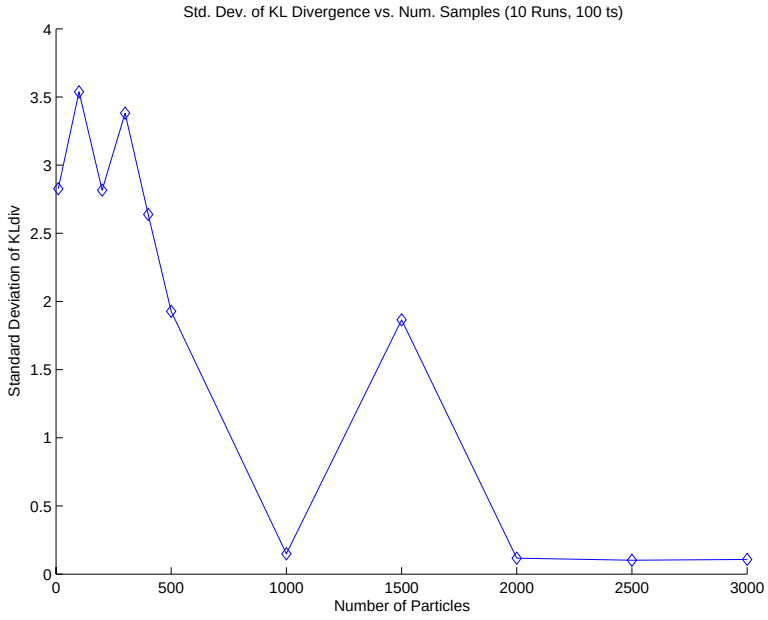


Figure 6-22: Variance, over 10 runs, of KL Divergence of the Grid-Sampled Posterior Distribution and the Kernelized Particle Filter Samples

is very low for the tests with 2000 or more samples.

In examining figure 6-21, we may also observe that the KL Divergence remains roughly constant as more than 2000 samples are used. This would support the idea that once a particle filter has many more samples than needed, performance improvements will be less significantly improved with additional samples.

6.2.2 KL Divergence vs. Resampling Threshold

Table 6.9: Parameters of the System

Pattern	Cosine
Number of Time-steps	100
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	1/4
Angle Process Noise	0
Spin-rate Process Noise	0

Table 6.10: Parameters of the Particle Filter

Number of Particles	1000
Resampling Threshold	0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1
Angle Regularization Variance	1/2
Spin-rate Regularization Variance	1/10

Table 6.9 describes the details of the system used to test the effects of the resampling threshold. All noise and system parameters were the same as those used in the tests with various numbers of samples.

Table 6.10 shows the parameters used for the particle filters. Each filter had 1000 samples, constant regularization variances, and resampling thresholds which varied between 0 and 1. The performance of the filter was evaluated at 1/20 increments of this resampling threshold. As explained in section 4.1, the system resamples when

the effective number of samples falls below the set threshold. If that threshold is 0, the particle filter will never resample, while if the threshold is 1, the filter resamples on every time-step.

The grid-sampled posterior for these system parameters is very similar to the one in presented in the previous section, figure 6-20. Since noise is minimal, the posteriors for each of these trials were effectively equivalent to the one previously shown. However the KL divergences, were, of course, still calculated with a posterior distribution that had identical observations as each particle filter.

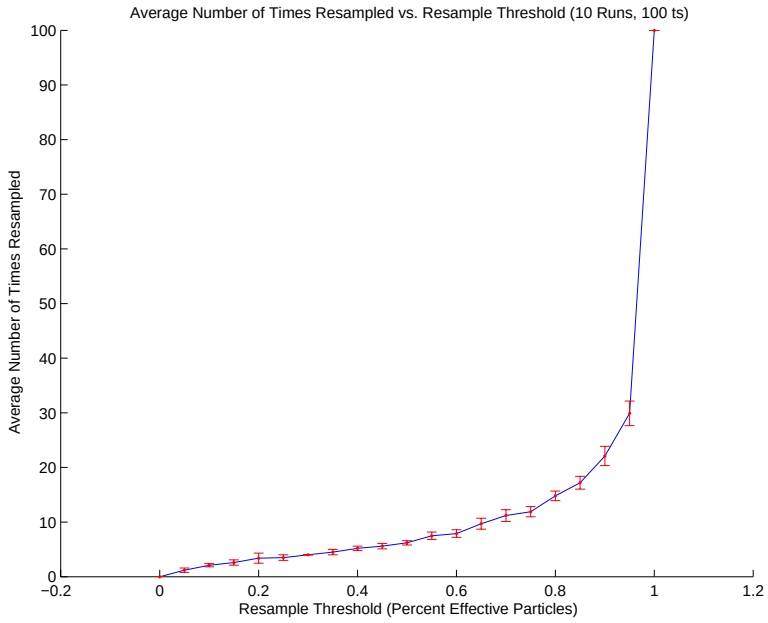


Figure 6-23: Average out of 10 runs of Number of Times Resampled (over 100 time-steps) vs. Resample Threshold. Vertical red bars represent the st. dev. of the averages.

Figure 6-23 shows the average number of times each of the particle filters resampled over the 100 time-steps. The frequency of resampling increases with the resample threshold, which is as one would expect. As the threshold increases, this increases the probability that the percent of effective particles will be below the threshold. When the threshold was 1, the particle filters did indeed sample on every time-step. Likewise, when the threshold was zero the filters never resampled.

Figure 6-24 is an example kernelized particle distribution for the 1000-sample filter with 0 as its resample threshold. Sample depletion is almost inevitable when the filter

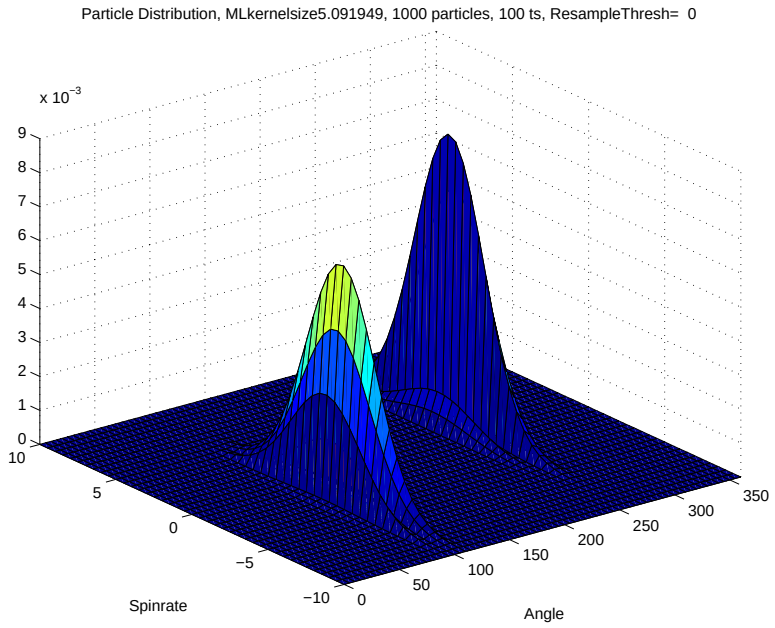


Figure 6-24: Kernelized Particle Distribution for a 1000-sample Particle Filter with the Resample Threshold at 0.

never resamples. The probability weights of the particles fall too quickly to zero. In addition, without resampling, there can be no regularization and the filter is limited by the granularity of its initial samples. However, despite these shortcomings, the average KL Divergence for the filters with resample thresholds of 0 was only a little higher than more moderate thresholds.

There was better performance in the filters with resample thresholds between 0.25 and 0.85. Figure 6-25 shows an example particle distribution for a particle filter with resample threshold of 3/5. This distribution managed to retain both peaks without much sample depletion or erroneous hypotheses.

Figure 6-26 is an example of the 1000-sample filters with the resample threshold at 1. This filter resampled on every time-step. This meant that there would be little smoothing over multiple time-steps since each particle was picked using a weight which was based only on its probability given the most current observation. We would expect the performance of this filter to be very poor since, unlike the others, it does not look at more than one observation to assist in choosing the highest likelihood samples.

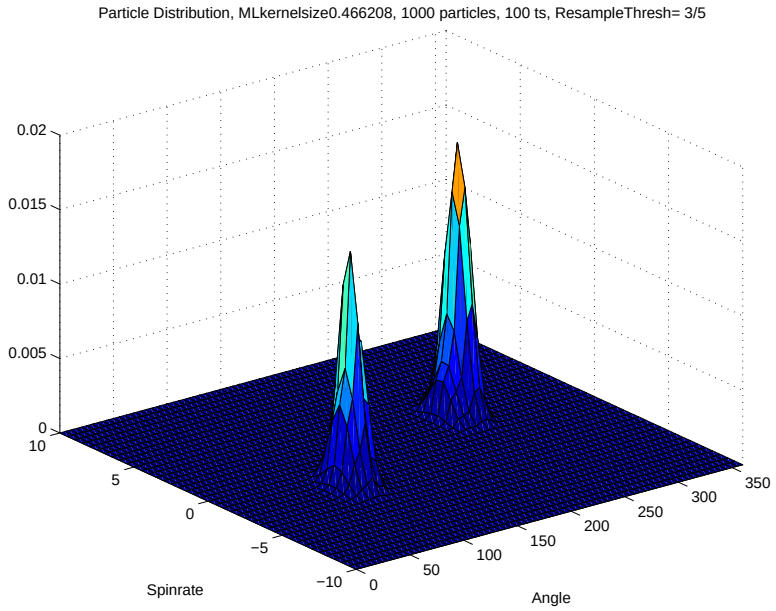


Figure 6-25: Kernelized Particle Distribution for a 1000-sample Particle Filter with the Resample Threshold at 3/5.

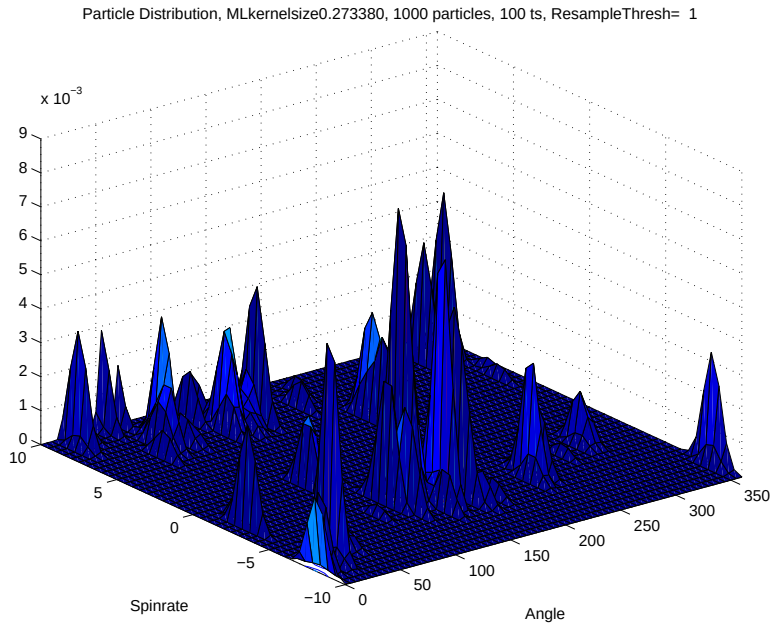


Figure 6-26: Kernelized Particle Distribution for a 1000-sample Particle Filter with the Resample Threshold at 1.

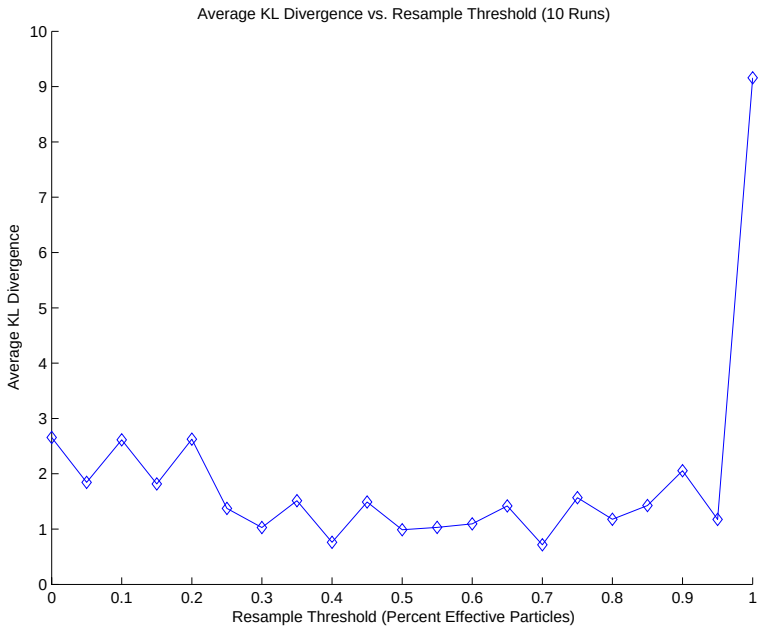


Figure 6-27: Average KL Divergence vs. Resample Threshold over 10 runs.

Figure 6-27 graphs the average KL divergence for each resample threshold. This graph shows all resample thresholds perform much better than when the threshold is 1. Even no resampling did better than resampling on every time-step. However, there seems to be no clear relationship beyond that since the average KL divergences are erratic.

Figure 6-28 shows the error bars which represent the variances of the average KL divergence at each resample threshold. Although it's once again hard to generalize, thresholds between 0.4 and 0.85 seem the most likely to have low KL divergences. The best of the tested thresholds, according to this graph would be 0.7.

6.2.3 KL Divergence vs. Regularization Variances

Table 6.11 describes the details of the system used to test the effects of regularization variances. All noise and system parameters were the same as for the tests with varying number of samples. Table 6.12 shows the parameters used for the particle filters. Each filter had 1000 samples, a constant resampling threshold, and varying angle and spin-rate regularization variances. The angle and spin-rate variances were

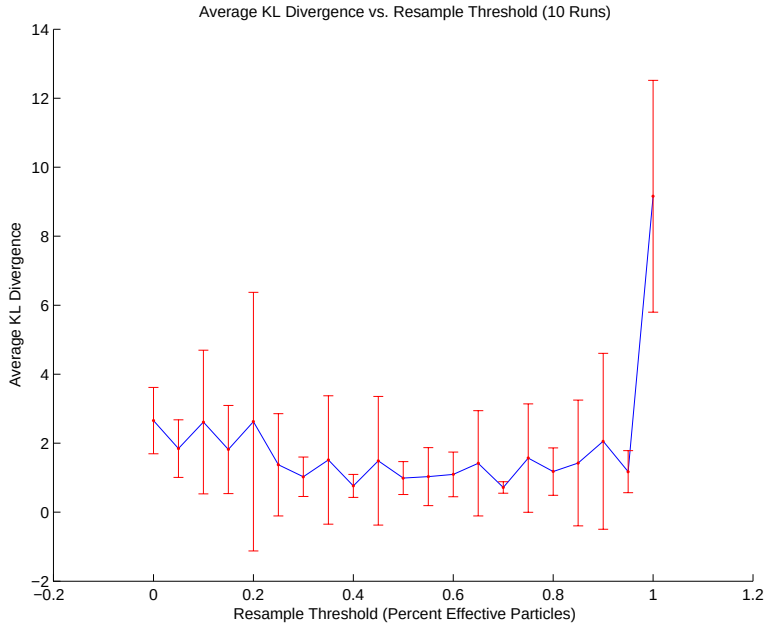


Figure 6-28: Average KL Divergence vs. Resample Threshold. The red vertical bars show the standard deviation of each average KL Divergence.

Table 6.11: Parameters of the System

Pattern	Cosine
Number of Time-steps	100
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	1/4
Angle Process Noise	0
Spin-rate Process Noise	0

manipulated simultaneously in the interest of condensing the results.

The grid-sampled posterior remains very similar to the one in figure 6-20, in the previous section, since the system parameters are unchanged. The only difference is the particular observations that were observed in each run.

Figure 6-29 shows the kernelized particle distribution for the particle filter with angle and spin-rate regularization variances both equal to zero. Without regularization the filter is susceptible to sample depletion, which is a major reason why at least one of the two correct peaks is lost in this figure.

Figure 6-30 graphs the kernelized particle distribution for the particle filter with

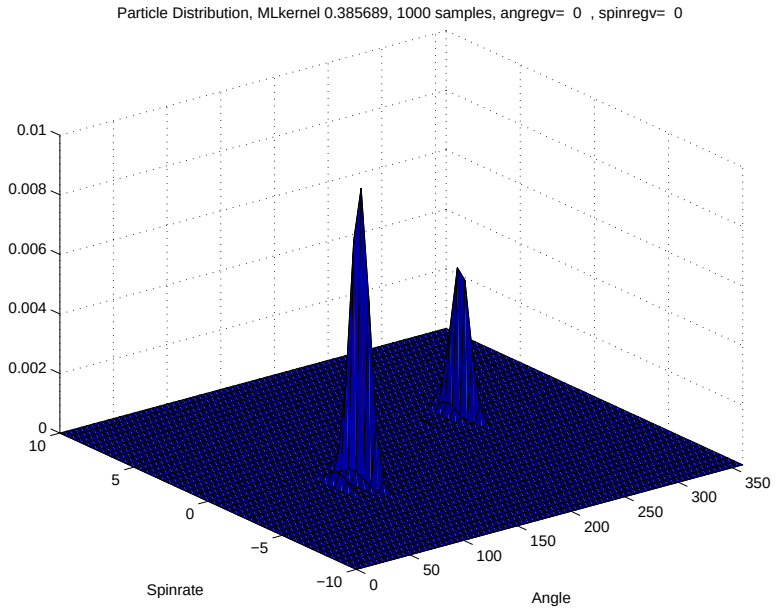


Figure 6-29: Kernelized Particle Distribution for Angle and Spin-rate Regularization Variances both equal to 0.

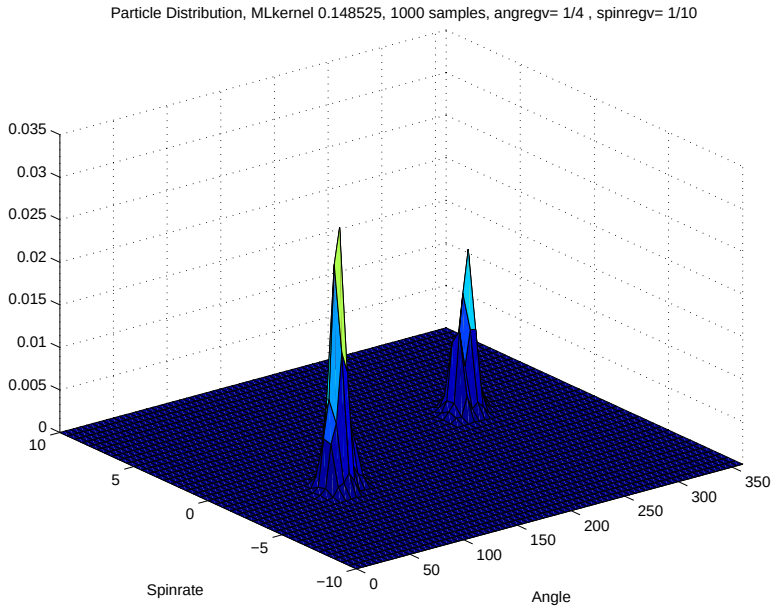


Figure 6-30: Kernelized Particle Distribution for Angle and Spin-rate Regularization Variances are 1/4 and 1/10, respectively.

Table 6.12: Parameters of the Particle Filter

Number of Particles	1000										
Resampling Threshold	5/6										
Angle Regularization Variance	0	.25	.5	.75	1	1.25	1.5	1.75	2	2.25	2.5
Spin-rate Regularization Variance	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

angle and spin-rate regularization variances of 1/4 and 1/10. This distribution is a large improvement over the one in figure 6-29 since is includes both peaks. On average the KL Divergence for the trials with these regularization variances was much less than the average KL Divergence for the trials with no regularization.

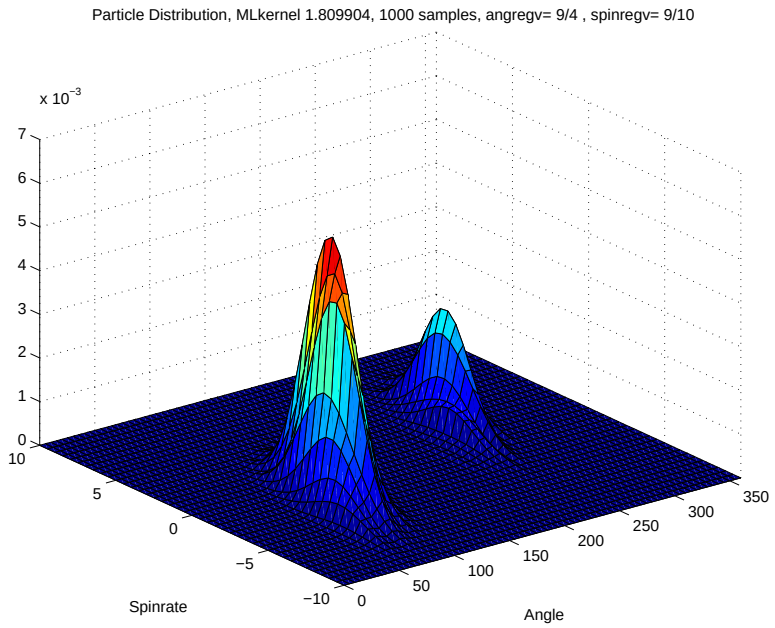


Figure 6-31: Kernelized Particle Distribution for Angle and Spin-rate Regularization Variances are 2.25 and 9/10, respectively.

Figure 6-31 shows the kernelized particle distribution for the particle filter with angle and spin-rate regularization variances of 2.25 and 9/10, respectively. This distribution is more spread out than the one in figure 6-30 since the gaussian resampling variances were much higher. On average the KL Divergence for the trials with these regularization variances was greater than the average KL Divergence for the trials with no regularization.

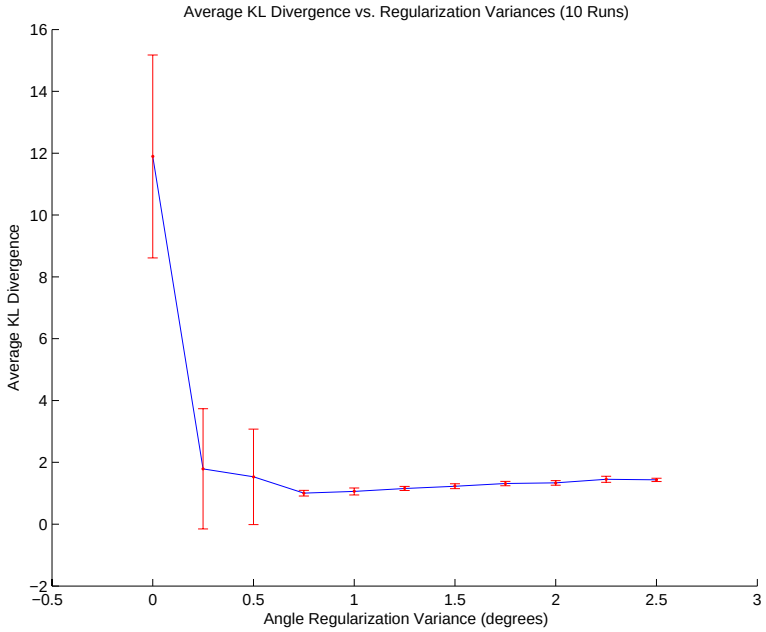


Figure 6-32: Average, over 10 runs, KL Divergence vs. angle regularization variance. The spin-rate regularization variance also changed according to table 6.12. The red vertical bars show the variance of each average KL Divergence.

Figure 6-32 shows the average KL Divergence of the sampled posterior with the kernelized particle filters with each of the angle regularization variances. The spin-rate regularization variance changed as well. The pairs of angle and spin-rate regularization variances were specified in table 6.12.

Clearly, there was a large improvement with small regularization variances versus no regularization. This is due to the particle filter's tendency towards sample depletion when regularization is not used. However, increased regularization variances eventually resulted in higher average KL Divergences. This is due to the artificially spread out peaks this encourages. The variance of the KL Divergence also increases since the width of the peaks is now highly related to when the last resampling time was.

6.2.4 KL Divergence vs. Noise Level

Table 6.13 describes the details of the system used to test the effects of noise on the performance of the particle filter. All system parameters were the same as those used

Table 6.13: Parameters of the System

Pattern	Cosine
Number of Time-steps	100
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1
Angle Process Noise	0
Spin-rate Process Noise	0

Table 6.14: Parameters of the Particle Filter

Number of Particles	1000
Resampling Threshold	5/6
Angle Regularization Variance	1/2
Spin-rate Regularization Variance	1/10

in previous tests. However, the variance of the gaussian angle observation noise varied between 0 and 20.

Table 6.14 shows the parameters used for the particle filters. Each filter had 1000 samples as well as constant regularization variances and resampling thresholds. Each particle filter was given the value of the variance of the measurement noise before beginning the task of filtering the observations. This is reasonable since the variance on the measurements is a detail of the sensor which can be estimated before filtering.

Figure 6-33 graphs the average KL divergence as the angle measurement noise varies between 0 and 1. The divergence is very high when there is no noise at all. The reason for this is that the posterior distribution has no uncertainty when there is no observation noise. The probability that a sample will be exactly the correct answer is impossibly small, so sample depletion is very likely. However, upon the addition of small amounts of noise, the filter's performance improves drastically.

However, when the noise increases above 0.2 the KL divergence again begins to increase. The finite samples of a particle filter are not very good at modeling large amounts of uncertainty. Because of their optimistic view of the uncertainty, particle filters are more sensitive to noisy observations.

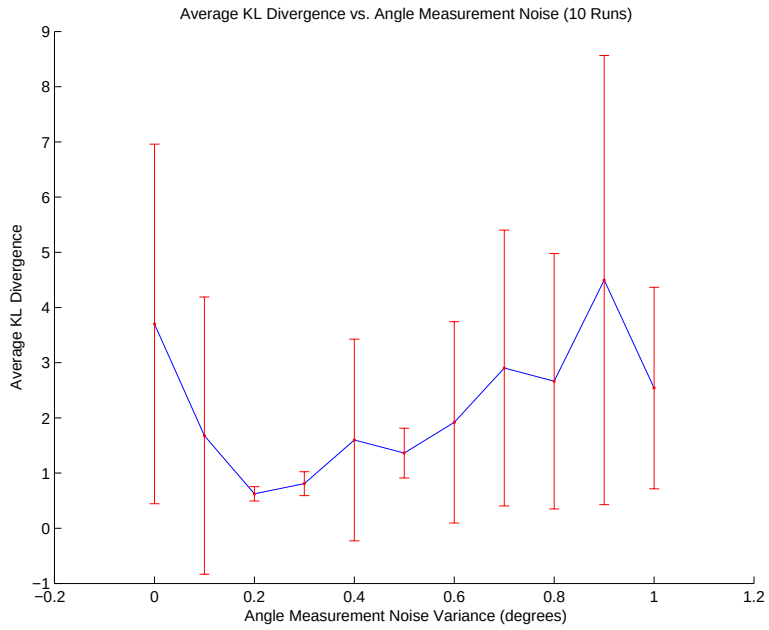


Figure 6-33: Average, over 10 runs, KL Divergence vs. Angle Measurement Noise. The red vertical bars show the variance of each average KL Divergence.

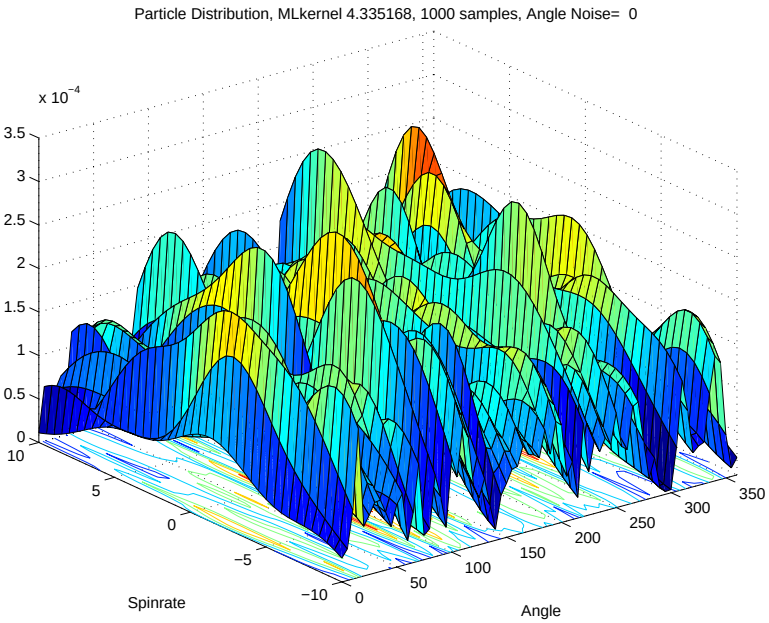


Figure 6-34: Kernelized Particle Distribution for the Particle Filter with the Angle Measurement Noise of 0.

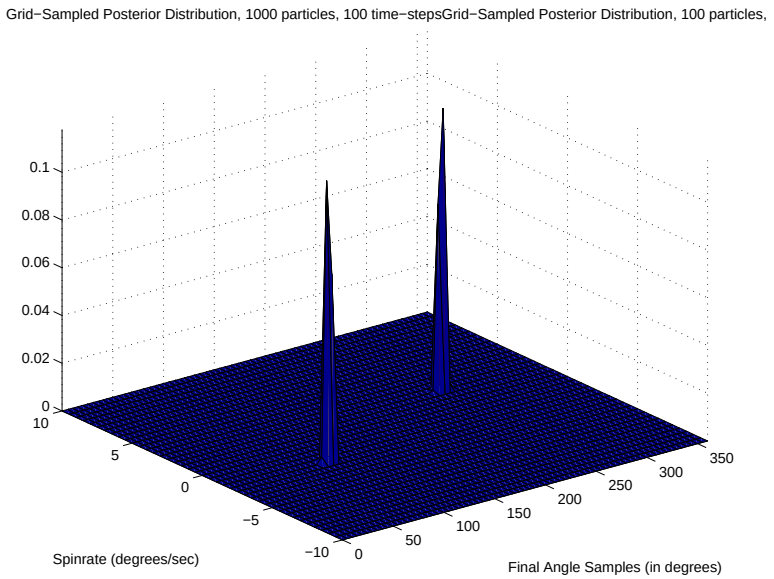


Figure 6-35: The Grid-Sampled Posterior Distribution for the Angle Measurement Noise of 0.

Figure 6-34 is an example particle distribution for the filter presented with observations with no noise. Since none of the samples were exactly right, they all seem equally bad. Hence it is very difficult for the filter to find the high probability regions with its limited number of samples, and the particle filter suffers sample depletion. The grid-sampled posterior distribution for the spinning discs system with no noise is shown in figure 6-35. While this distribution is very sharp it is not really correct since the grid is not infinitely fine.

Figure 6-36 shows an example kernelized particle distribution when the noise was 0.2. This distribution is an improvement since there is at least one peak for one of the hypotheses. The filter has not lost both of the peaks in the posterior distribution as the lower noise filter had. Figure 6-37 is the corresponding grid-sampled posterior distribution. This distribution is still quite sharp since there is little noise and 100 time-steps of observations.

An example particle distribution for the noisiest system parameters is shown in figure 6-38. The particle filter did pretty well considering the low signal to noise ratio when the noise variance is equal to the maximum value of the pattern. Although there are a number of peaks in the generally correct areas. However the erroneous peaks

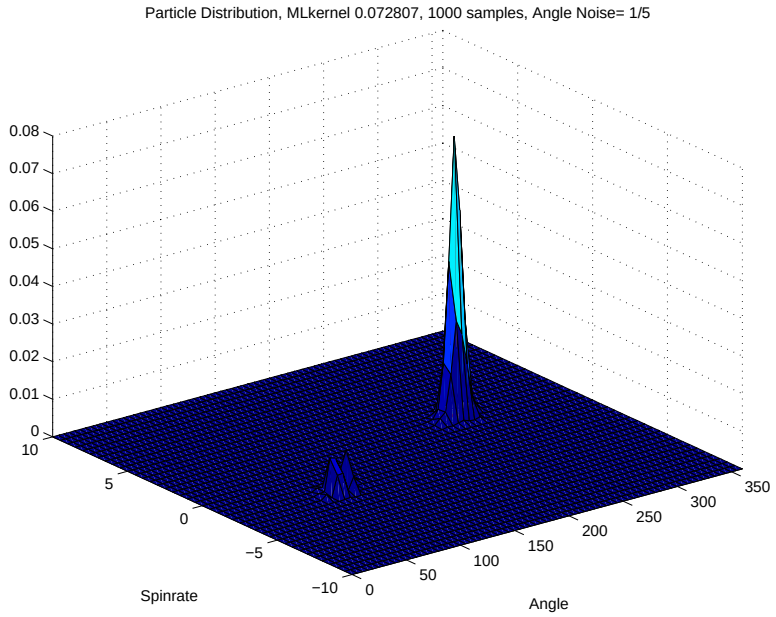


Figure 6-36: Kernelized Particle Distribution for the Particle Filter with the Angle Measurement Noise of 1/5.

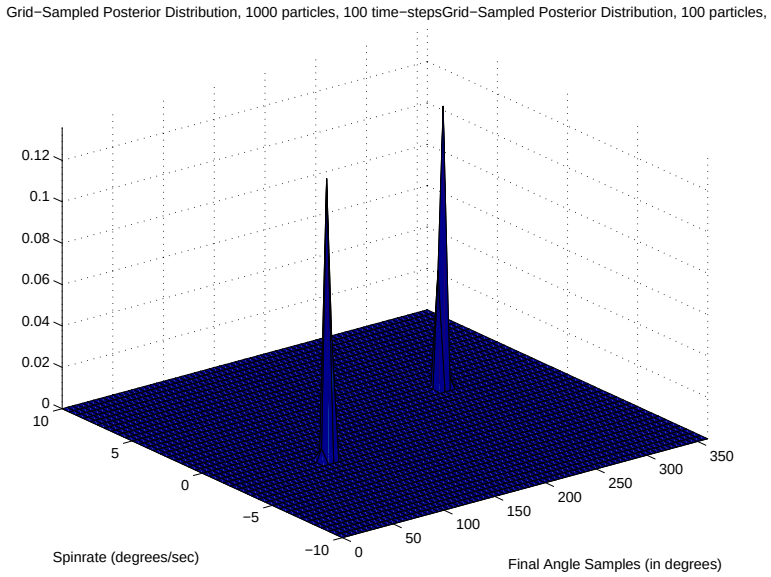


Figure 6-37: The Grid-Sampled Posterior Distribution for the Angle Measurement Noise of 1/5.

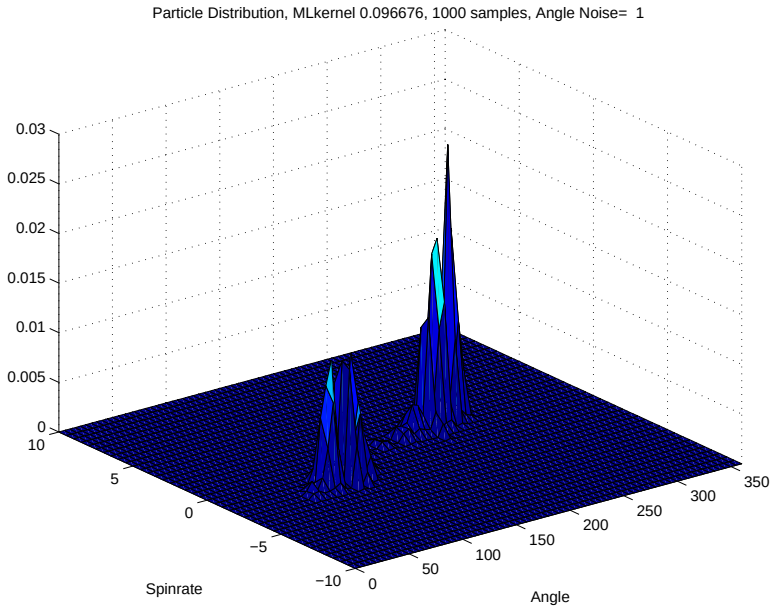


Figure 6-38: Kernelized Particle Distribution for the Particle Filter with the Angle Observation Noise of 1.

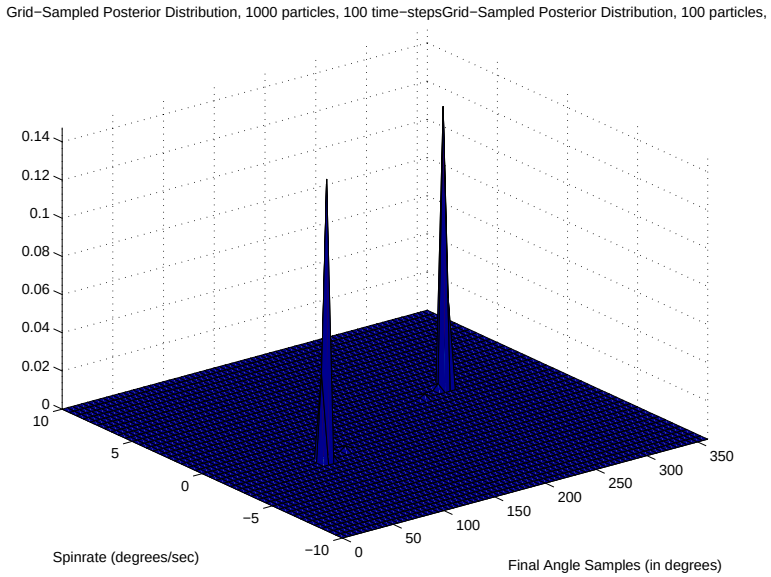


Figure 6-39: The Grid-Sampled Posterior Distribution for the Angle Measurement Noise of 1.

are due to the noise sensitivity of the limited samples. The posterior distribution shown in figure 6-39 is still very sharp, due to the combination of 100 observations.

6.2.5 KL Divergence vs. Number of Time-Steps

The last factor to affect the performance of the particle filters is the number of time-steps of observations. Sampling issues again show themselves, since with very few observations there will be too much uncertainty for the finite number of samples to model. In addition, as the number of observations increases, if the posterior distribution becomes too sharp the particle filter may have difficulty modeling it. However, regularization of the particles, also called gaussian resampling, may avoid the threat of sample depletion in cases with a large sequence of observations.

Table 6.15: Parameters of the System

Pattern	Cosine
Number of Time-steps	1,5,10,25,50,75,100,150,200
Initial Angle	65 degrees
Spin-rate	2 degrees/sec
Angle Measurement Noise	1/4
Angle Process Noise	0
Spin-rate Process Noise	0

Table 6.16: Parameters of the Particle Filter

Number of Particles	1000
Resampling Threshold	5/6
Angle Regularization Variance	1/2
Spin-rate Regularization Variance	1/10

Table 6.15 describes the details of the system used to test the relationship of the number of time-steps and the performance of the 1000-sample particle filter. All other system parameters were the same as those used in the previous tests.

Table 6.16 shows the parameters used for each of particle filters. Each filter had 1000 samples, angle regularization variance of 1/2, spin-rate regularization variance

of 1/10, and resampling thresholds of 5/6.

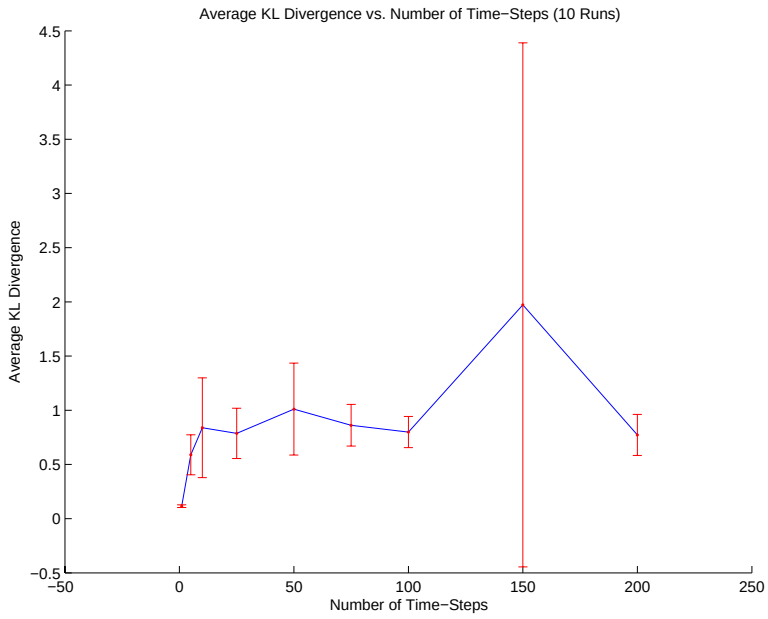


Figure 6-40: KL Divergence of the Grid-Sampled Posteriors vs the Particle Distributions for Various Numbers of Time-steps. The red bars represent the standard deviations of the averages.

Figure 6-40 shows the average KL Divergence between the distribution of the grid-based filter and the kernelized distribution of the particle filter. The KL Divergence is highest for the 150-step particle filter.

Figure 6-41 shows the maximum likelihood Parzen density estimate of the particle distribution after 10 time-steps of observations. This broad distribution is actually somewhat similar to the posterior grid-sampled distribution shown in figure 6-42. With only 10 observations there is still quite a bit of uncertainty in the angle and especially in the spin-rate. The particles modeled this uncertainty quite well. The average KL divergence for the 10-time-step filters, in figure 6-40, is comparatively low.

Figure 6-43 is the kernelized particle distribution for a particle filter after seeing 25 time-steps of observations. Clearly the distribution is much narrower. Figure 6-44 shows the posterior grid-based filter distribution with much thinner peaks. However the average KL divergence after 25 steps was higher than for 10 time-steps.

The 200-time-step particle filter did not see a drop in average KL divergence.

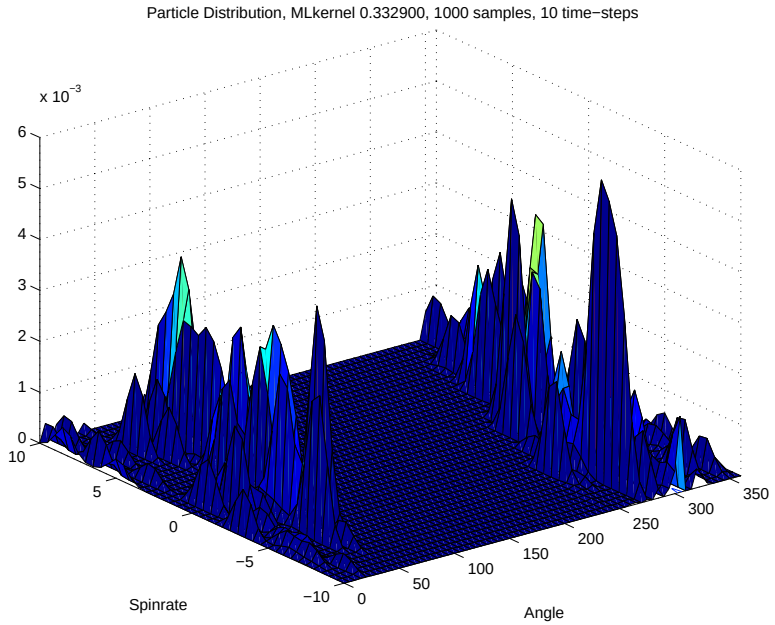


Figure 6-41: Kernelized Particle Distribution for the Particle Filter run for 10 time-steps.

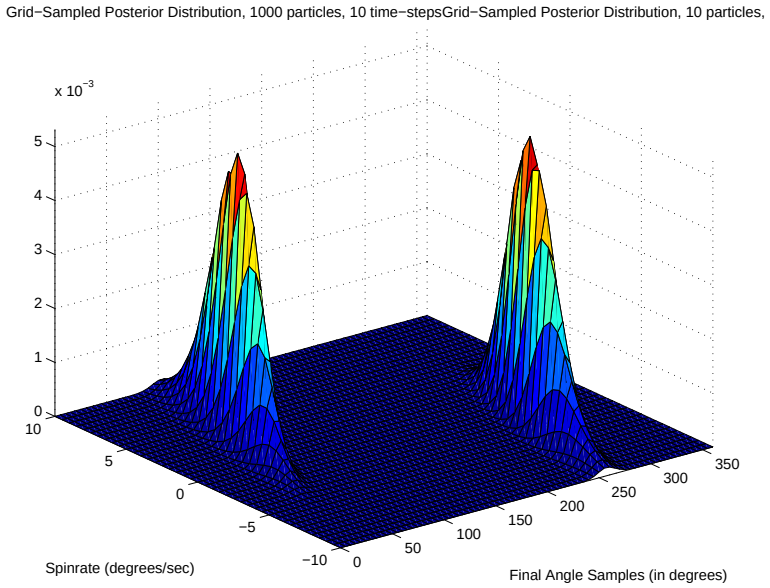


Figure 6-42: The Grid-Sampled Posterior Distribution for 10 time-steps of observations.

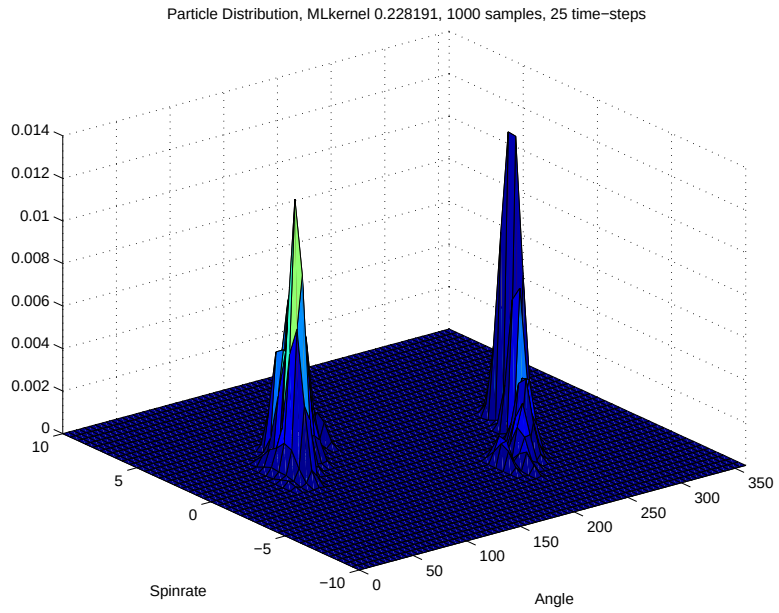


Figure 6-43: Kernelized Particle Distribution for the Particle Filter run for 25 time-steps.

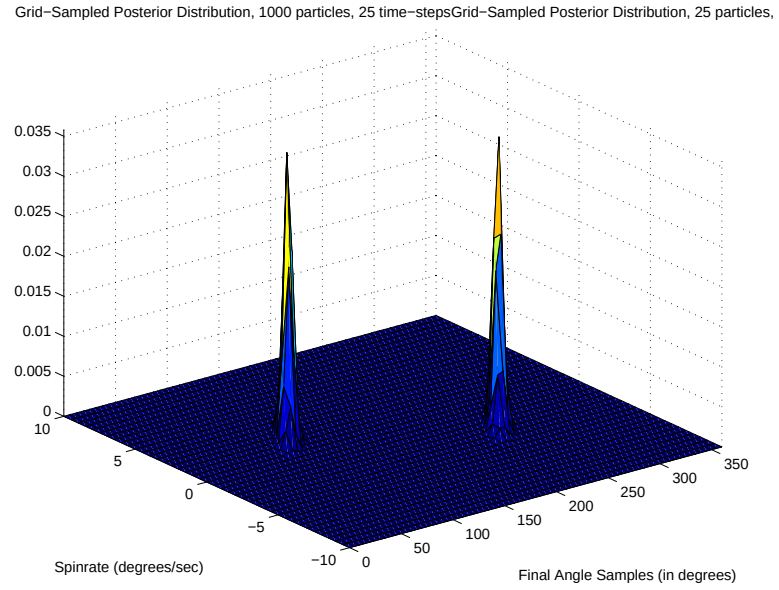


Figure 6-44: The Grid-Sampled Posterior Distribution for 25 time-steps of observations.

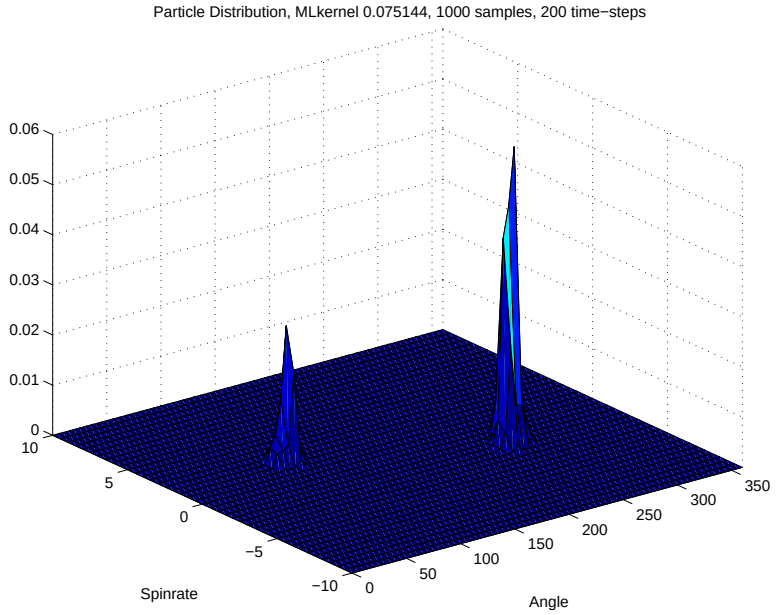


Figure 6-45: Kernelized Particle Distribution for the Particle Filter run for 200 time-steps.

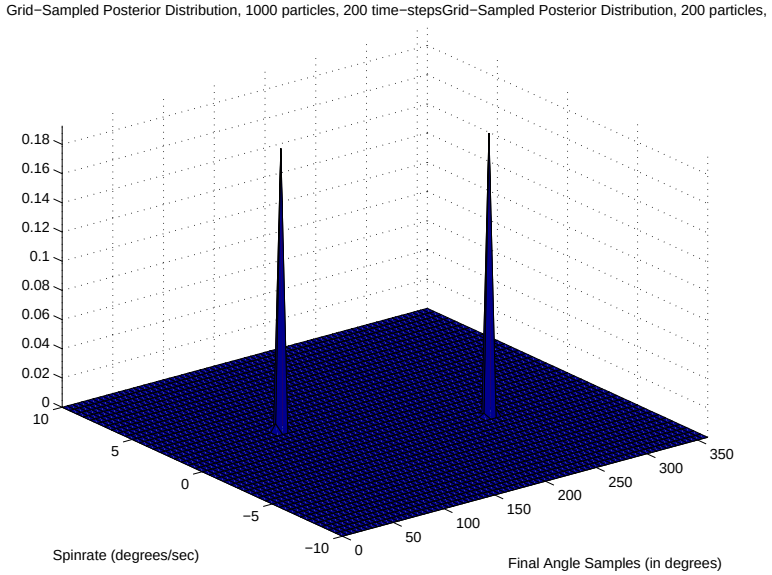


Figure 6-46: The Grid-Sampled Posterior Distribution for 200 time-steps of observations.

Despite having the most observations, the reported KL divergence was still comparable to those of the particle filters with fewer time-steps. This could be because the regularization variances limited how sharp the posterior particle distribution could be. Another cause could be that the true posterior is too sharp for the grid used in the grid-based filter and our supposedly more accurate distribution is not so accurate.

The comparison of the cosine particle filter to the grid-based filter showed the sensitivity of the particle filter to its number of samples, resampling threshold, regularization variances, noise level, and number of time-steps. The overall results show that the particle filter can give reasonable estimates of the posterior density when there is enough regularization, a moderate resampling threshold, and sufficiently many particles. In addition, the noise level must be non-trivial for reasonable performance without sample depletion. Finally, as the number of time-steps becomes very large, the particle distribution of the particle filter may be pessimistic due to the regularization variances.

6.3 Discrimination With More Challenging Patterns

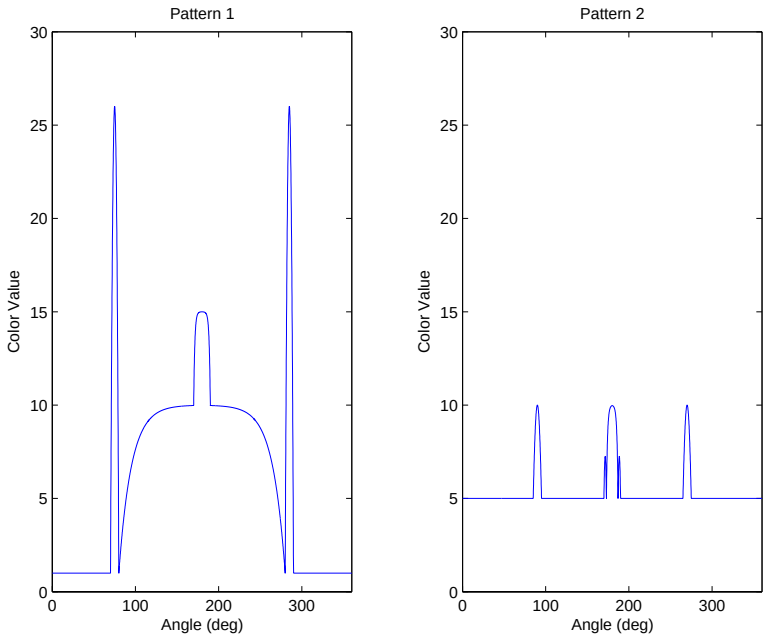


Figure 6-47: The Two More Peaked Patterns Used for Discrimination

Figure 6-47 shows the two more challenging patterns which were used to test the particle filter for its ability to discriminate among patterns. These patterns are more challenging because some observations correspond to many possible states, such as an observation of 5, while others may map to very few possible states, such as an observation of 26. This poses a sample depletion problem to a particle filter.

At a given time the particle filter may be modeling a very broad distribution. For patterns 1 and 2, this would be the case when the estimator is observing the pattern in any of the flat regions. If the filter then receives an observation that should narrow that distribution down to a very narrow peak, the likelihood of samples being in that narrow peak will be very small. This leads to sample depletion. The phenomenon of sample depletion when the posterior distribution was extremely narrow compared to the immediately preceding prior distribution occurred in the cosine particle filter when the noise level was too small.

The results of testing the particle filter on these new, more peaked, patterns

displayed a similar sensitivity to the sharpness of the posterior distribution. Although the particle filter often discriminated the correct pattern from the observations, its ability to estimate an accurate posterior distribution for angle and spin-rate was very limited.

Table 6.17: Parameters of the System

Pattern	Pattern 1
Number of Time-steps	20
Initial Angle	65 degrees
Spin-rate	1/2 degrees/sec
Angle Measurement Noise	5
Angle Process Noise	0
Spin-rate Process Noise	0

Table 6.18: Parameters of the Particle Filter

Number of Particles	1000
Resampling Threshold	1/4
Angle Regularization Variance	1/2
Spin-rate Regularization Variance	1/10

Table 6.17 shows the system parameters used to test discrimination of the particle filter. Table 6.18 details the parameters used for the filter. Now the state estimates were augmented by an estimate the pattern. The grid-based filter used for comparison used the same grid of angle and spin-rate for each pattern to estimate the full posterior density of pattern, angle, and spin-rate.

Figure 6-48 displays the patterns estimates of the particle filter over the 20 time-steps. The particles quickly converged on pattern 1 as the correct pattern. Despite the very noisy observations, the posterior probability of pattern 2 given the many sequential observations is very low. However, the estimation of the correct angle and spin-rate of the samples performed very poorly after 20 time-steps.

Figure 6-49 is the kernelized particle distribution for the pattern 1 particles after 1 time-step. The distribution is a good approximation of the true posterior, as shown

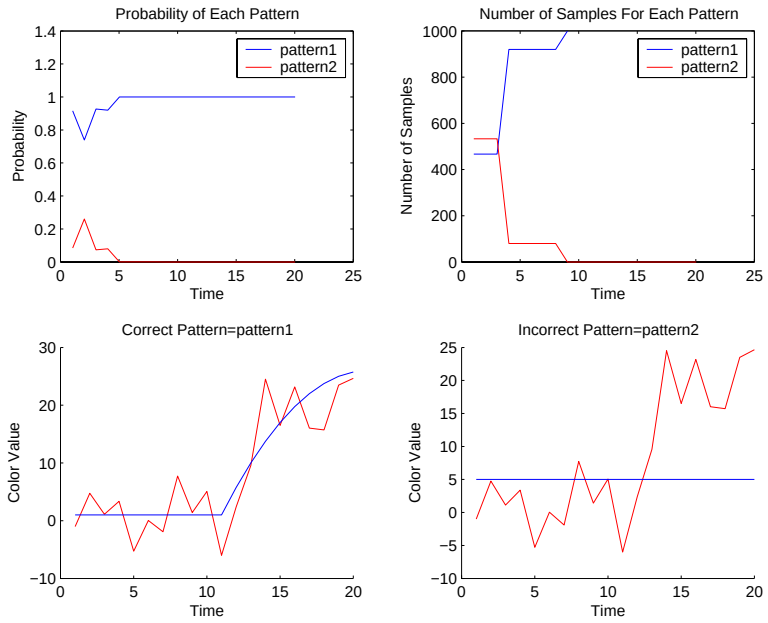


Figure 6-48: The Pattern Estimate of the 1000-sample Particle Filter Over 20 Time-Steps

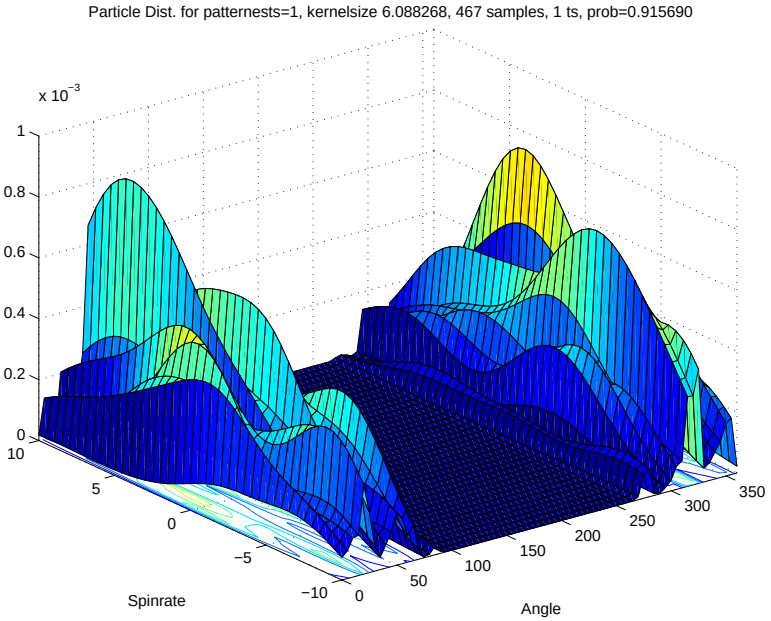


Figure 6-49: The Kernelized Particle Distribution for the Pattern 1 Samples after 1 time-step

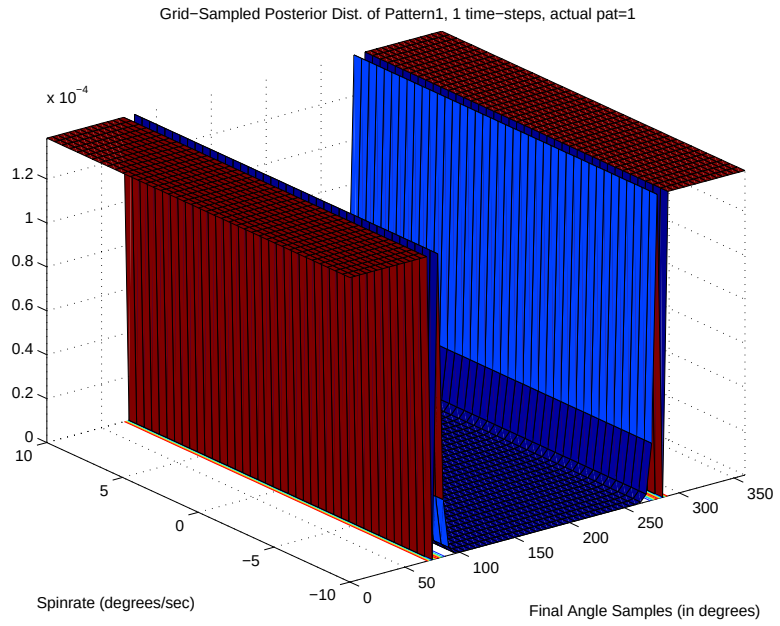


Figure 6-50: The Grid-Sampled Posterior Distribution for the Pattern 1 Samples after 1 time-step

in the fine grid-based filter for pattern 1 samples in figure 6-50. In general the KL divergence of the particle distribution is quite low after only one time-step. However, as time progresses and the posterior distribution narrows, the particle filter does a worse job of approximating the posterior distribution.

After 20 time-steps, the posterior distribution favors the angles 50 and 250 and the spin-rates +10 and -10. The particle distribution in figure 6-51 corresponds reasonably well to the posterior distribution, in figure 6-52, which has even sharper peaks for the favored samples. The observations are misleading here, since these favored states are not the true ones. The true states in the grid-sampled distribution still have non-zero probability weights, although their probability weights are very low. In the particle distribution the samples in the correct area have already been resampled away. Unfortunately, after 10 time-steps, the correct answers are not the most likely solutions given the observations.

Figure 6-53 is the particle distribution after 20 time-steps. New observations have eliminated what was before the favored states. However, since the correct states were already resampled, the particles converged on a local maximum in the wrong area.

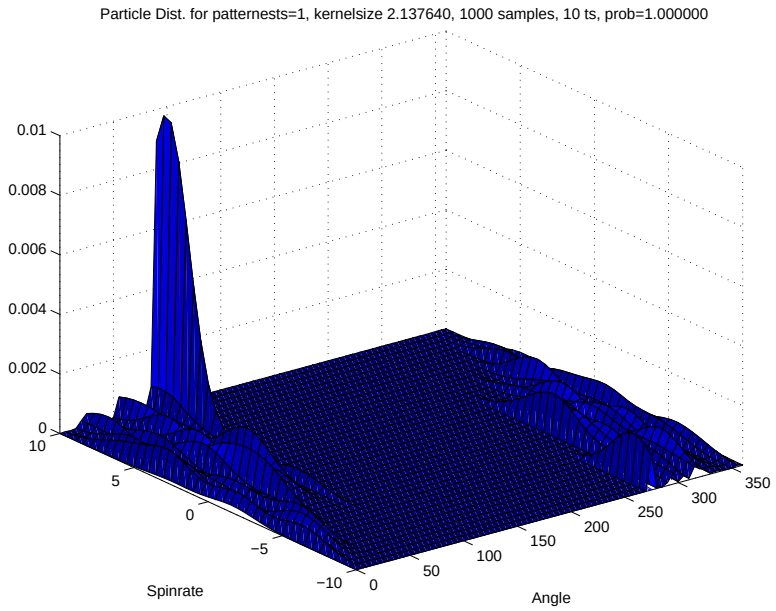


Figure 6-51: The Kernelized Particle Distribution for the Pattern 1 Samples after 10 time-steps

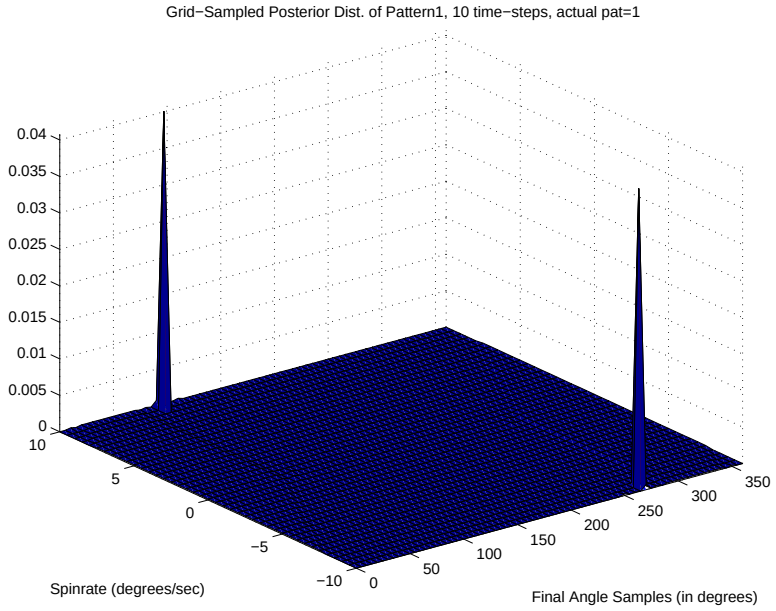


Figure 6-52: The Grid-Sampled Posterior Distribution for the Pattern 1 Samples after 10 time-steps

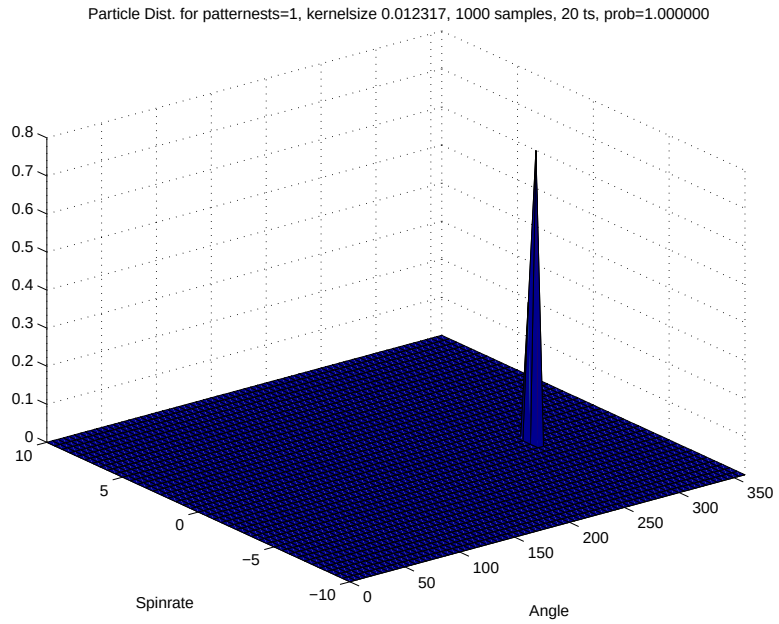


Figure 6-53: The Kernelized Particle Distribution for the Pattern 1 Samples after 20 time-steps

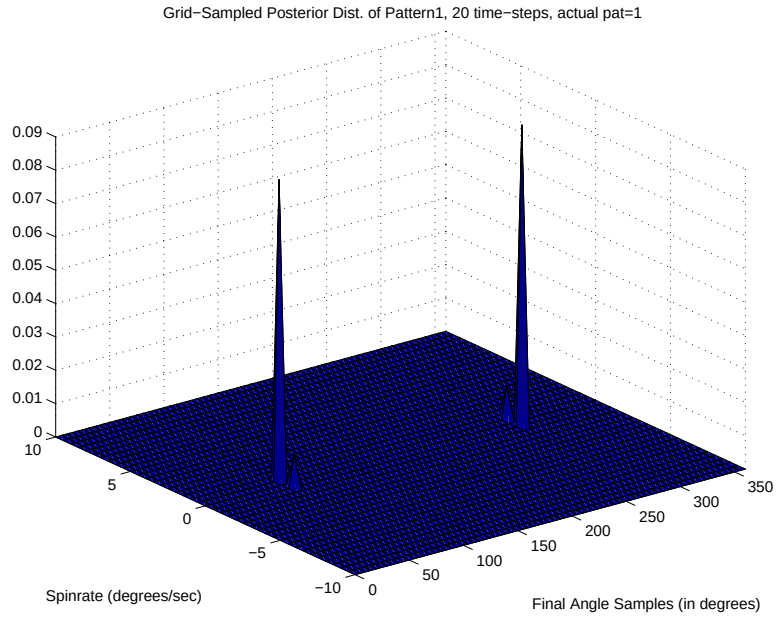


Figure 6-54: The Grid-Sampled Posterior Distribution for the Pattern 1 Samples after 20 time-steps

Figure 6-54 shows the grid-sampled posterior distribution. The gridded distribution has converged to the correct regions since there is no resampling in the grid-based filter. In general the KL divergence of the particle filter, after 20 time-steps, was very large due to sample depletion. These more peaked patterns can be misleading, causing sample depletion.

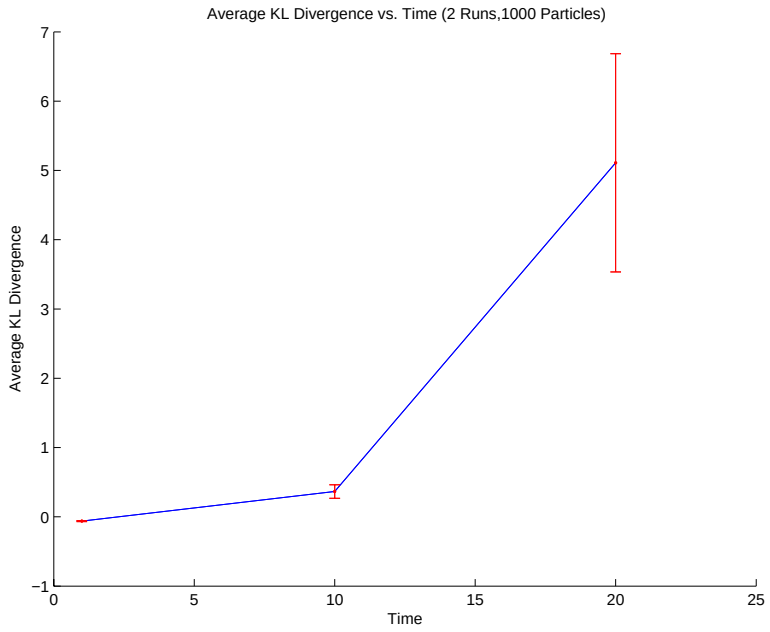


Figure 6-55: Average KL Divergence for the Grid-Sampled Posterior 1000-sample Particle Filter Particle Distribution for 1, 10, and 20 time-steps. The red bars represent the standard deviation of the averages.

Figure 6-55 shows the average KL Divergence of grid-based filter's distribution and the kernelized particle distribution at 1, 10, and 20 time-steps. The KL divergence was very small at first, but then jumped after 20 time-steps. This jump is due to the sample depletion which almost always occurred between 10 and 20 time-steps.

Overall, the particle filter did demonstrate its ability as a classifier for these more challenging patterns. It classified the patterns correctly most of the time after only a few time-steps. However, the particle filter did a very poor job of estimating the distribution of over angle and spin-rate with the highly peaked patterns. The sharp peaks in the patterns created situations where new observations were not too likely

given previous observations, either because the prior distribution was too broad or was misleading, which usually led to sample depletion and large errors.

Chapter 7

Conclusions and Further Study

The goal of this thesis was to perform filtering and estimation on observations of spinning discs using a particle filter. The performance of the particle filter was evaluated in both cases where the measurement model was linear and nonlinear.

In the linear case, the particle filter was shown to approximate the least squares solution given by the Kalman filter. Cases in which the particle filter performed badly were those in which the noise level was extremely low or the number of particles was too small.

The particle filter was also tested on the cosine spinning discs pattern. The sensitivity of the filter to the number of particles, resampling threshold, regularization variances, measurement noise level, and number of time-steps was examined. The key parameters were the number of particles and noise level. As long as there were enough particles, non-zero noise levels, and the remaining parameters were at moderate values, the particle filter performed well.

Finally, the particle filter was tested for its ability to discriminate between two more challenging patterns. It was found that although, the correct pattern was usually chosen, the filter performed poorly at estimating the correct angle and spin-rate of the more peaked patterns. This result shows the deep sensitivity of the filter to the patterns themselves. The regions in the peaked patterns with very high slopes, following regions of very low slope, resulted in posterior distributions that were much sharper than the distribution for the previous time-step. Sample depletion was a

major problem with these types of patterns. From these results, we would predict that similar sample depletion scenarios would be common in patterns with both low and high frequency regions.

Future work on particle filtering for spinning discs could focus on the effects of modeling the particle as Gaussian distributions. The technique of Gaussian Blurring, which James Williamson suggested, was not used in these tests due to time constraints. Gaussian Blurring refers to modeling the particles each as a kernelized Gaussian distribution, rather than a delta function, when calculating the likelihood of the sample. By modeling the uncertainty with which the particles approximate the posterior density, this technique may improve the robustness of the particle filter in the situations that often caused sample depletion.

In addition, there is current work, by Rudolf van der Merwe and Eric Wan, studying a new type of Monte Carlo sampling filter that uses samples to model multi-modal Gaussian mixtures as distributions.[17] The propagation of these modes through the state transition equation uses the Unscented Transform of the UKF. Since the Gaussian mixture-model is a more complete representation of a distribution than discrete samples, perhaps this extension of the idea of particle filter will be less susceptible to sample depletion.

In conclusion, Particle Filtering is a promising technique for the approximate inference of multi-modal posterior distributions resulting from nonlinear systems. However, the we must continue to study the limitations of the particle filter and improve sampling performance to make Particle Filtering a solution to more challenging nonlinear problems.

Bibliography

[1] Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions On Signal Processing*, 50(2):174–189, February 2002.

[2] Yaakov Bar-Shalom, X.-Rong Li, and Thiagalingam Kirbarajan. *Estimation with Applications To Tracking and Navigation*. John Wiley & Sons, Inc., New York, 2001.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[4] Dan Crisan and Arnaud Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions On Signal Processing*, 50(3):736–746, March 2002.

[5] Arnaund Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer, New York, 2001.

[6] Dieter Fox. Kld-sampling: Adaptive particle filters. In *Advances in Neural Information Processing Systems 14*. MIT Press, 2001.

[7] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, London, England, 1970.

[8] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new approach for filtering nonlinear systems. *Proceedings of the American Control Conference*, pages 1628–1632, 1995.

[9] S. J. Julier and J. H. Uhlmann. A new extension of the kalman filter to nonlinear systems. *Proceedings of the AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[10] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering (Series D.)*, 82:35–45, 1960.

[11] R.E. Kalman and R.S. Bucy. New results in linear filtering and prediction theory. *Transactions of the ASME Journal of Basic Engineering (Series D.)*, 83:95–108, 1961.

[12] Peter S. Maybeck. *Stochastic models, estimation and control*. Academic Press, New York, 1982.

[13] Peter S. Maybeck. *Kalman filtering : concepts, design and extensions ; a tutorial presented at the 51st MORS, Air Force Institute of Technology, September 27 and 28, 1983*. Wright-Patterson Air Force Base, OH : Institute of Technology, 1983., 1983.

[14] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, second edition, 1997.

[15] Stuart Russell and Peter Norvig Editors. *Artificial Intelligence A Modern Approach*. Prentice Hall Series in Artificial Intelligence. Pearson Education, Inc., Upper Saddle River, New Jersey, second edition, 2003.

[16] Rudolf van der Merwe and Eric Wan. The unscented kalman filter for nonlinear estimation. *Proceedings of Symposium 200 on Adaptive Systems for Signal Processing, Communication and Control(AS-SPCC)*, 2000.

[17] Rudolf van der Merwe and Eric Wan. Sigma-point kalman filters for probabilistic inference in dynamic state-space models. *Proceedings of the Workshop on Advances in Machine Learning*, June 2003.